Towards the origin of Lyme Borreliosis

Vollmer, Stephanie

Award date:
2010

Awarding institution:
University of Bath

Towards the origin of Lyme Borreliosis

By Stephanie Anne Vollmer

A thesis submitted for the degree of Doctor of Philosophy

University of Bath
Department of Biology and Biochemistry

August 2010

Author Signature ……

Abstract

Lyme borreliosis (LB) is the most frequent vector-borne disease in the Northern Hemisphere. It is a complex bacterial zoonosis involving vertebrate hosts and hard ticks of the genus *Ixodes*. The causative agents, bacteria of the LB group of spirochaetes, form a species complex comprising 17 named species. As is the case for most microbial pathogens, epidemiological and ecological studies require appropriate genotyping. Although the use of single loci may provide rapid results, there are serious disadvantages, in particular when inferring evolutionary relationships or geographic population structure. A novel multilocus sequence analysis (MLSA) system of the LB group spirochaetes has been developed based on housekeeping genes to overcome these problems. Here, the system is optimized and tested using extracted spirochaetal DNA directly from ticks and then utilized to obtain insights into the migration and spread of individual species as well as to investigate the evolutionary origins of the species complex. Species belonging to the LB group of spirochetes display different patterns and levels of host specialisation which makes this an ideal system to study the impact of host associations on spread of zoonotic tick-borne diseases. For example, *Borrelia valaisiana* and *B. garinii* are transmitted exclusively by birds while *B. afzelii* is transmitted by rodents. I demonstrate that the migration of the LB species is dependent on, and limited by, the migration of their respective hosts. I also show the presence of *B. afzelii* strains in England and, through the use of the MLSA scheme, demonstrate that the strains are highly structured. A close evolutionary relationship between *B. afzelii* and its rodent host species is shown. Furthermore, through phylogenetic analyses, some evidence of a coevolutionary relationship between the LB group species and their major group of vector species, the *Ixodes persulcatus* species complex, is presented and a Eurasian origin for the species group is suggested.

Table of Contents

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations

µl	Microlitre
aLRT	Approximate likelihood ratios
BIONJ	Biological neighbour-joining analysis
BLAST	Basic local alignment search tool
CC	Clonal complex
DLV	Double locus varient
dN/dS	Ratio of non-synonymous to synonymous
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
Freq	Frequency
F_{ST}	Fixation statistic
GTR	General time reversable
HKY	Hasegawa, Kishino and Yano
IGS	Intergenic spacer
LB	Lyme borreliosis
LGM	Last glacial maximum
ml	Milliliter
ML	Maximum Likelihood
MLSA	Multilocus sequence analysis
MLST	Multilocus sequence typing
n	Number of samples
NCBI	National centre for biotechnology information
NH_4OH	Aqueous amonia
NJ	Neighbour-joining analysis
OSP	Outer surface protein
π	Mean nucleotide pairwise distance
P	Probability
PCR	Polymerase chain reaction
p-distance	Mean nucleotide pairwise distance
p-value	probablility value
θ	Estimated population mutation rate
RNA	Ribonucleic acid
rrf-rrl	5S-23S intergenic spacer
rRNA	Ribosomal ribonucleic acid
SDW	Sterile distilled water
SH-like	Shimodaira-Hasegawa-like
SLV	Single locus varient
SNP	Single nucleotide polymorphism
SPR	Subtree pruning and regrafting
ST	Sequence type
τ	Estimated mutational steps since expansion
TAE	Tris-acetate-Ethylenediamine-tetraacetic acid
UV	Ultra violet
X^2	Chi squared test

Acknowledgements

Firstly, I would like to thank my late supervisor, Klaus Kurtenbach whose enthusiasm for the subject was infectious and I am exceptionally grateful for his encouragement and the opportunities he created for me whether it was to attend conferences and workshops or taking part in field work projects. I also owe a great debt of gratitude to Gabi Margos and Ed Feil who stepped in to fill the role of supervisor after Klaus' death. They have both given up large amounts of their time to ensure I could complete my thesis. Thanks to all the members of 3 South, lab 1.09 past and present who have supported me throughout my PhD and provided me with many happy memories of my time at Bath. Special thanks to Kevin Balbi for his patient assistance with bioinformatics programmes and helping me learn to write scripts in Perl. I would also like to thank Santiago Castillo-Ramirez for his guidance on phylogenetic analyses.

A number of people have assisted me in my work by providing ticks and cultured *Borrelia* samples for testing using the multilocus scheme. Special thanks to Antra Bormane for ticks from Latvia, Marianne James for Scottish ticks, Frederik Seelig for German ticks, Andrew Dobson and Sarah Randolph for additional English ticks and Muriel Cornet for cultured *Borrelia* samples from France. Thanks to Shun Yue Chan and Laura Brett for screening of British ticks and David Aanensen for his assistance with the MLST website.

Finally, I would like to thank my family for their continuing support while I have been completing my PhD. I would especially like to thank, my husband, Chris Bannister for his endless patience, advice and support throughout my studies at Bath University.

Chapter 1: Introduction

1.1 Taxonomy

Borrelia are part of the bacterial group of spirochaetes. Members of this group have a distinctive helical shape and are highly motile, with seven to eleven periplasmic flagellae at each end of the cylindrical cell (Goldstein *et al.*, 1996, Goldstein *et al.*, 1994). The genus *Borrelia* forms a monophyletic clade of species that is split into three distinct clusters (Barbour and Hayes, 1986, Takano *et al.*, 2010). One cluster incorporates the relapsing fever spirochaetes, another includes the newly described relapsing-fever like spirochetes, and the third cluster comprises of the Lyme Borreliosis (LB) group of spirochaetes. The LB group presently consists of 17 named species and several putative species yet to be named (Margos *et al.*, 2009, Rudenko *et al.*, 2009a, Rudenko *et al.*, 2009b). However, only a small proportion of these are known to cause disease in humans.

Borrelia burgdorferi was first isolated in 1981 from the hard tick, *Ixodes scapularis* collected in New York State in the United States (Burgdorfer *et al.*, 1982). The isolate was subsequently identified as a new species of the genus *Borrelia* and was named *B. burgdorferi* after its discoverer Willy Burgdorfer (Johnsen *et al.*, 1984). In the following years many closely related isolates were cultured across the northern hemisphere from a broad range of reservoir hosts and tick species as well as samples from patients with different disease symptoms (Barbour *et al.*, 1985, Bissett and Hill, 1987, Stalhammar-Carlemalm *et al.*, 1990, Wilske *et al.*, 1988). It became apparent through molecular studies that the *B. burgdorferi* isolates were genetically and phenotypically diverse (Postic *et al.*, 1990). So the group was redefined as a species complex known as the *B. burgdorferi* sensu lato (s.l.) and phenotypically distinct groups of isolates were referred to as different genospecies (Baranton *et al.*, 1992, Canica *et al.*, 1993, Fukunaga *et al.*, 1996b). The genospecies *B. burgdorferi*, first isolated in 1981, was then referred to as *B. burgdorferi* sensu stricto (s.s.) to differentiate it from the *B. burgdorferi* s.l. species complex (Baranton *et al.*, 1992, Kawabata *et al.*, 1993). However, due to clear evidence showing that genospecies formed distinct species groups, more recently the group has been referred to as the LB group of spirochaetes (which conforms with the naming of the relapsing-fever group of spirochaetes) with members referred to as species instead of genospecies (Kurtenbach *et al.*, 2010) and this is how the group will be referred to throughout this thesis.

1.2 Epidemiology

LB is considered an emerging zoonotic disease and cases of Lyme disease are increasing in several areas of the globe including the United States and Europe (CDC, 2008, HPA, 2010b, Lindgren and Jaenson, 2006). Four members of the LB species group have been found to regularly cause disease symptoms in humans. These are *B. afzelii*, *B. burgdorferi*, *B. garinii* and *B. spielmanii* (Busch *et al.*, 1996, Fingerle *et al.*, 2008, Nadelman *et al.*, 1996, Strle *et al.*, 1996). They also tend to cause different disease symptoms, although this is a general pattern and not a rule as reviewed by Wang and colleagues (1999), *B. afzelii* is most frequently linked with skin manifestations (Canica *et al.*, 1993), *B. garinii* with neuroborreliosis (Ornstein *et al.*, 2001, Rijpkema *et al.*, 1997, Ruzic-Sabljic *et al.*, 2001) and *B. burgdorferi* s.s. with arthritic symptoms (Ornstein *et al.*, 2001, van Dam, 2002, van Dam *et al.*, 1993) and *B. spielmanii* is associated solely with erythema migrans (EM) (Fingerle *et al.*, 2008, Foldvari *et al.*, 2005). EM (also known as the bulls eye rash) is possibly the most common symptom associated with LB in humans but also varies in rate of occurrence between the different species, occurring most commonly with *B. afzelii* and *B. spielmanii* infections (reviewed by (Wang *et al.*, 1999b). Laboratory diagnosis of LB usually follows an internationally recommended two-step approach, using antibody screening tests as a first stage, followed by immuno-blotting (western blotting) of samples that gave reactive results in the screening tests (HPA, 2010a, Wilske, 2003, Wilske *et al.*, 2000).

The most highly pathogenic form of Lyme disease is caused by a newly defined species called *B. bavariensis* . This species was previously a subgroup of *B. garinii* strains known as *ospA* serotype 4 strains. Due to the fact they were genetically, ecologically and clinically distinct from other *B. garinii* strains, they were reclassified as a new species (Margos *et al.*, 2009). *B. bavariensis* is associated most often with neuroborreliosis but is rarely identified in ticks suggesting it is likely to be highly invasive (Marconi *et al.*, 1999, van Dam *et al.*, 1997).

Several other LB group species have occasionally been associated with disease symptoms in humans. *B. valaisiana* has been associated with EM in one patient (Rijpkema *et al.*, 1997) and *B. lusitaniae* has been identified in two patients suffering from Lyme disease (Collares-Pereira *et al.*, 2004, de Carvalho *et al.*, 2008). *B. bissettii* has been identified in patients in Europe but not in the US (Rudenko *et al.*, 2008, Rudenko *et al.*, 2009c). This may be related to vector species that transmit *B. bissettii* on the different continents, as some tick species are more likely to bite humans than others.

1.3 Vectors and Hosts

There are four principal tick vectors of the disease-causing LB group spirochaetes these are *Ixodes persulcatus*, *I. ricinus*, *I. scapularis* and *I. pacificus* (Piesman and Gern, 2004). The different species are often linked with specific vectors as illustrated in Table 1.1. However, the full vector range of LB species is much larger and is not fully understood. Eisen and Lane (2002) suggest a total of 42 ticks species have been found harbouring LB infections, but it is important to realise that they may not necessarily be able to transmit the spirochaetes on to other reservoir hosts. Ticks must be tested experimentally to identify whether they are able to transmit the infection (Eisen and Lane, 2002a). It is known that hard ticks of the *Ixodes persulcatus* species complex (also referred to as the *I. ricinus* or *I. scapularis* species complex) form the main group of tick species involved in the transmission of LB spirochaetes (Eisen and Lane, 2002a). All species within this group that have been experimentally tested are able to transmit the spirochaetes (Dolan *et al.*, 1998, Eisen and Lane, 2002a, Gern and Rais, 1996, Lane *et al.*, 1994, Nakao and Sato, 1996). However, the transmission of LB is not exclusive to this complex and several *Ixodes* species outside of the group have also been shown to transmit LB species as well. For example, *I. uriae*, the sea bird tick is known to transmit *B. garinii* between sea birds (Olsen *et al.*, 1993).

Table 1.1 The major vector species of the LB group of spirochaetes in different global regions modified from Masuzawa (2004).

Borrelia	North America	Europe	Russia	China	Japan
B. americana	*I. minor*				
B. andersonii	*I. dentatus*				
B. bissettii	*I. spinipalpis* *I. pacificus*				
B. burgdorferi	*I. scapularis* *I. pacificus*				
B. californiensis	Unknown				
B. carolinensis	*I. minor*				
B. afzelii		*I. ricinus*	*I. persucatus*		
B. bavariensis		*I. ricinus*	*I. persucatus*		
B. garinii	*I. uriae*	*I. ricinus* *I. uriae*	*I. persucatus*		
B. lusitaniae		*I. ricinus*			
B. spielmanii		*I. ricinus*			
B. valaisiana		*I. ricinus*			*I. columnae*
B. sinica				*I. ovatus*	
B. yangzte				*I. granulatus*	
B. japonica					*I. ovatus*
B. tanukii					*I. tanuki*
B. turdi					*I. turdi*

Hard ticks have four developmental stages in their life cycle: egg, larvae, nymph and adult (Figure 1.1). Their life cycle takes two to five years to complete depending of

climatic conditions (Randolph *et al.*, 2002). At each feeding stage (i.e. larva, nymph and female adult) the ticks must quest for a host and if they find one they feed for one to seven days depending on species, developmental stage and environmental conditions (Balashov, 1972). After feeding the tick detaches from its host, falls to the ground and enters into a diapause period of up to one year. During this period the ticks digests the blood meal and moults to the next developmental stage or, in the case of mated females, they prepare to lay between 1,000 and 10,000 eggs (Hillyard, 1996). A large proportion of ticks are lost at each developmental stage and it has been suggested that this is due to failure in finding a host (Gray *et al.*, 1992, Ostfeld *et al.*, 2001). Abundance and seasonal appearance (phenology) of ticks are crucial factors in the ecology of LB (Kurtenbach *et al.*, 2006) and the absence of the vector essentially means the absence of bacteria and human borreliosis. Figure 1.1 shows that most Ixodid ticks feed on a wide range of host species (Estrada-Pena *et al.*, 2005, Talleklint and Jaenson, 1997). While immature ticks (larvae, nymphs) feed on small animals such as rodents and birds, adult female ticks prefer to feed on large animals such as deer (Figure 1.1). Adult male ticks quest for a host but do not feed. As each female adult tick may give rise to many eggs, deer act as the reproductive host of the ticks (Stafford *et al.*, 2003, Wilson *et al.*, 1985). It has been shown that the absence of deer means absence of tick species that vector LB, such as *B. scapularis* (Wilson, 1986, Wilson *et al.*, 1984) and hence absence of LB. Some studies have produced promising results in reducing tick numbers by excluding deer from areas or inhibiting ticks from feeding on deer by applying pesticides to wild deer populations (Daniels and Fish, 1995, Schulze *et al.*, 2008).

It is believed that in the US the spatial demographic expansion of deer populations is leading to expansions in the demographic population size of LB spirocheate species (Diuk-Wasser *et al.*, 2006, Lastavica *et al.*, 1989). In Britain there has been an increase in reported Lyme disease cases over the past 15 years (HPA, 2010b) and data from Ward (2005) reveals that populations of Roe deer (*Capreolus capreolus*) have expanded in many parts of the UK since 1972 (Figure 1.2). A similar situation seems to be occurring in the mid-west and east coast of the US where there has been a continuous increase in Lyme disease cases and massive population growth and range expansion of deer (Diuk-Wasser *et al.*, 2006, Lastavica *et al.*, 1989) reviewed by Spielman and colleagues (1985). This has been linked to habitat increase in the form of reforestation that occurred in the US in the mid 20th century (Spielman *et al.*, 1985).

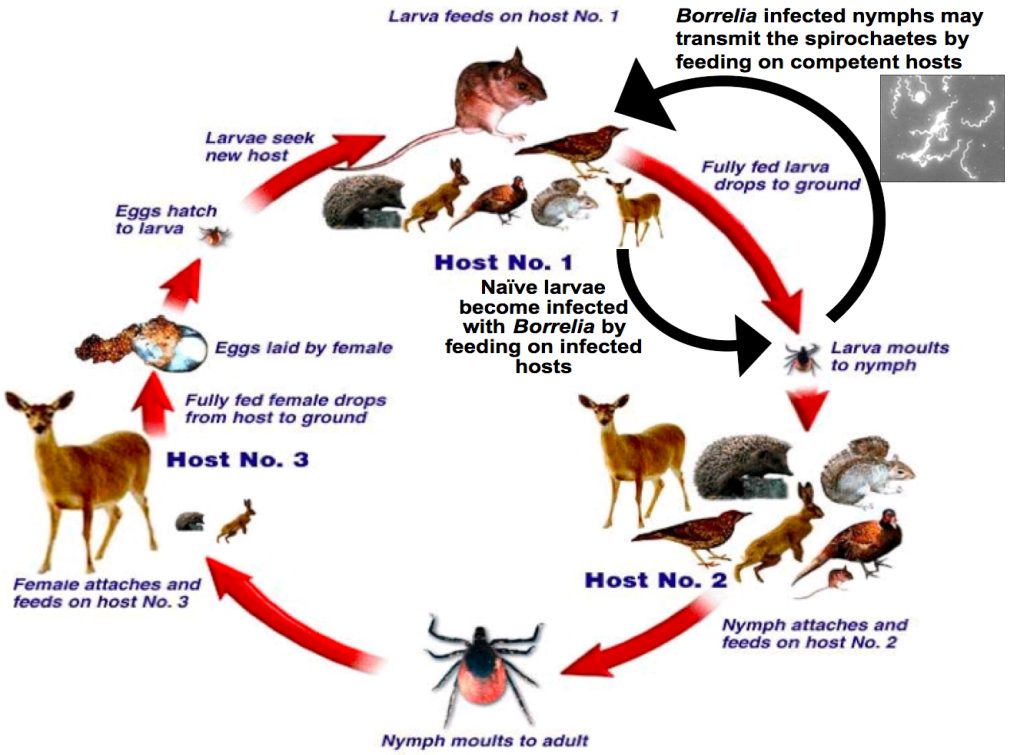


Figure 1.1 Tick life cycle showing their relationship with host species. The relative size of the animals approximates their significance as hosts for that stage of the tick life cycle. Each tick stage moults during a phase of diapause which can take 1 year (Designed by B. Kaye).

While deer are an essential factor for the presence of many *Ixodes* tick species, they are not directly an important factor in the life cycle of the spirochaetes. This is for two reasons, (i) deer are not competent to transmit spirochetes and, therefore, do not act as a reservoir host (Kurtenbach *et al.*, 2002a) and (ii) spirochaetes are rarely transmitted transovarially from adult ticks to their offspring (Kurtenbach *et al.*, 1995, Randolph and Craine, 1995, Zhioua *et al.*, 1994). The black arrows in Figure 1.1 illustrate the typical transmission cycle of LB spirochaetes where naïve larvae acquire spirochaetes by feeding on an infected animal host. Then, after moulting to the nymphal stage, they may transmit the infection to a naïve animal host during the blood meal (Eisen and Lane, 2002a). Humans are effectively a dead end host for the spirochaetes as they are rarely in forest environments to be part of the tick feeding cycle and play a role in forward transmission.

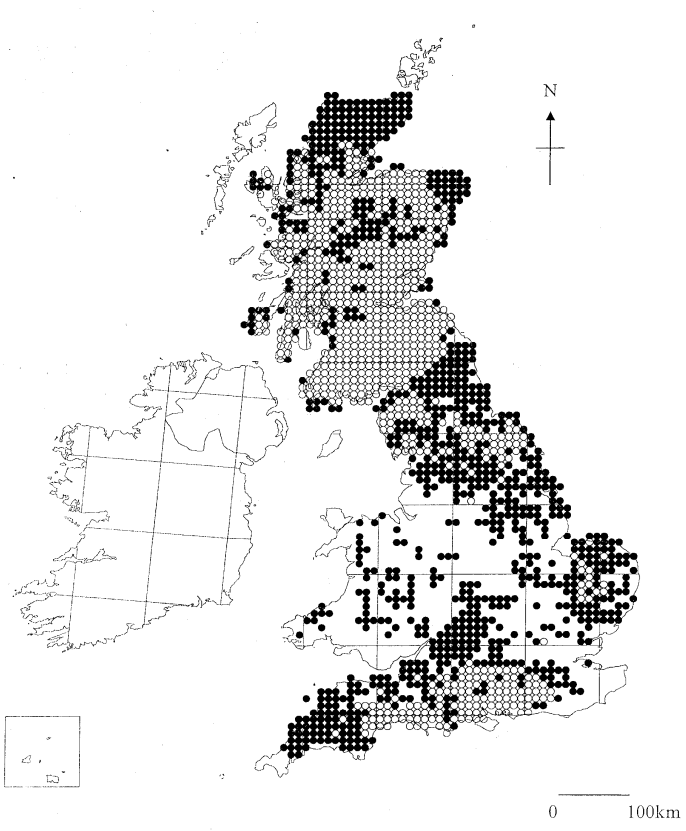


Figure 1.2. Map of Great Britain showing the expansion of Roe deer (*Capreolus capreolus*) from 1972 to 2002 (Ward, 2005). White circles represent areas where deer thought to be present in 1972 and are also present today, while black circles represent areas where deer were believed to be present in 2002 but were not present in 1972.

While *I. ricinus*, *I. persulcatus* and *I. scapularis* are considered generalist feeders, the majority of LB group species are host specialists, only able to be transmitted between ticks by certain host types. For example, *B. afzelii* is transmitted by rodents and some mammalian insectivores, while *B. garinii* and *B. valaisiana* are only able to be transmitted by avian species (Hanincova *et al.*, 2003a, Hanincova *et al.*, 2003b, Kurtenbach *et al.*, 2002a, Kurtenbach *et al.*, 2001). LB species also appear to be transmitted more frequently by certain host species within their host range. For example, it appears that song thrushes (*Turdus philomelos*) and blackbirds (*Turdus merula*) make up 99 % of the host species range transmitting *B. garinii* and *B. valaisiana* in forests in central Europe (Dubska *et al.*, 2009, Taragel'ova *et al.*, 2008). In England pheasants also appear to be an important host (Kurtenbach *et al.*, 1998). One reason that these bird species are more common hosts may be because they are ground-feeding, i.e. they spend a greater amount of time foraging on the forest floor where they come into contact with questing ticks. Whether there are other contributing factors to this phenomenon is currently unknown. A mechanism conferring this host association has been related to the spirochaete's ability to evade the host's innate immune system. More specifically, it has been shown that the interactions of a group of surface proteins found in *Borrelia* called complement regulator-acquiring surface proteins (CRASPs) (Bykowski *et al.*, 2008), inhibit the actions of components of the hosts complement system (Kurtenbach *et al.*, 2002a, Kurtenbach *et al.*, 2002c). These host

associations mean that the ecology of local sites, such as the structure of the reservoir host community, greatly influences the prevalence of different LB group species (Kurtenbach *et al.*, 1998) as well as influencing spread and migration of the species (Vollmer *et al.*, 2010).

1.4 Geographic Distribution

LB is the most common vector borne disease of the temperate zones of the northern hemisphere (Steere *et al.*, 2004). However, the LB species are not evenly distributed across this region as shown in Figure 1.3. Some species such as *B. afzelii* and *B. garinii* have been found throughout Eurasia (Baranton *et al.*, 1992, Masuzawa, 2004, Mediannikov *et al.*, 2005) while other species have a highly localised distribution, such as *B. lusitaniae* which is mainly found in regions around the Mediterranean Sea (Huegli *et al.*, 2002). Host specialisms of the LB spirochaetes are likely to influence the global distribution of different spirochaetal species. Although not indicated by the map, *B. garinii* possibly has the broadest distribution of all the LB group spirochaetes. Not only is it found in forest regions across Eurasia, it is also maintained in sea bird colonies by the tick vector, *I. uriae*. This means it is also found in many far-reaching sites including arctic regions and colonies off the east coast of the US (Duneau *et al.*, 2008, Smith *et al.*, 2006). It is astonishing given the wide distribution of *B. garinii* and the apparent overlap of terrestrial and seabird cycles in Europe (Comstedt *et al.*, 2009) that in the United States, *B. garinii* has not spread into inland areas.

The relative abundance of species throughout different regions varies probably due to ecological conditions. For example *B. burgdorferi* is highly prevalent in the mid-west and the east coast of the US but, although found in Europe, it is relatively rare and has not been identified in Asia (Masuzawa, 2004). *B. afzelii* is often the most common species identified in mainland Europe but is relatively rare in England (Vollmer *et al.*, 2010).

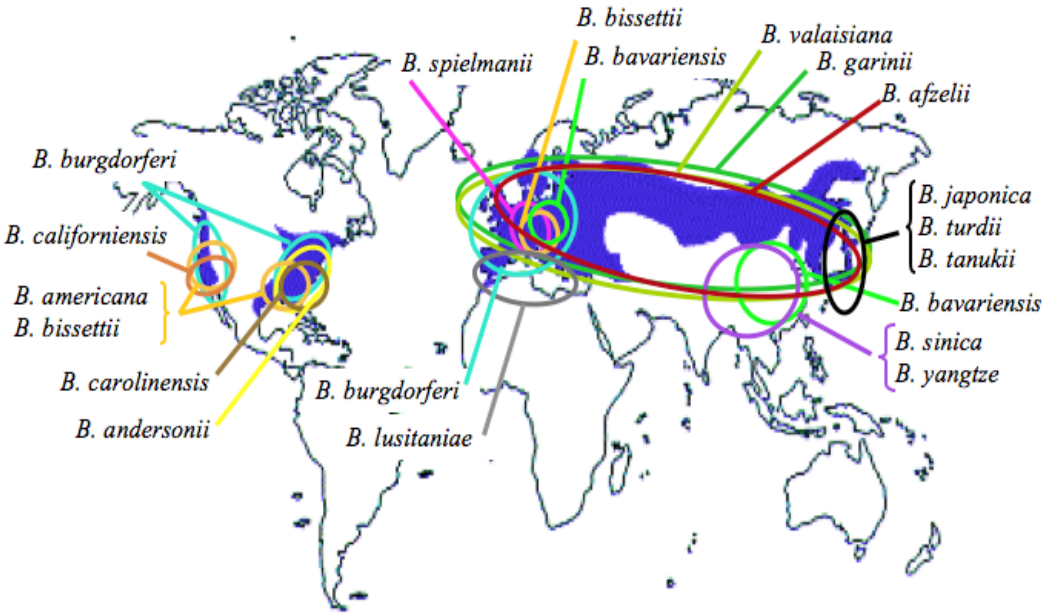


Figure 1.3. Map showing the global distribution of the LB species. The shaded blue areas show the distribution of Lyme disease cases.

1.5 Genome Organisation

To date there are six whole genome sequences available for four different LB species. These are *B. garinii* strain PBi (recently renamed *B. bavariensis* (Glockner *et al.*, 2004, Margos *et al.*, 2009), *B. burgdorferi* strain B31 (Fraser et al. 1997), *B. burgdorferi* strain ZS7, *B. afzelii* strain Pko (Glockner *et al.*, 2006), *B. afzelii* strain ACA and *B. spielmanii* A14S. There are also four relapsing fever spirochaete genomes available and many more LB species genome projects ongoing.

Borrelia species have unique genomes in that they have one linear chromosome of just over 900 kb but also have a number of linear and circular plasmids. They have the largest number of extra chromosomal elements found in any bacterium. In *B. burgdorferi* B31 an extra 610 kb is spread across 21 plasmids, 9 circular and 12 linear as shown in Table 1.2 (Casjens *et al.*, 2000, Fraser *et al.*, 1997, Glockner *et al.*, 2006). Many of these plasmids are found in all natural isolates of a particular LB group species and they may also contain essential genes (Barbour, 1993). For example, plasmid located genes include those that encode CRASPs that have been demonstrated to bind host complement regulator factor H or factor H-like molecules, protecting LB spirochaetes from host complement-mediated killing (Kraiczy *et al.*, 2001, Stevenson and Miller, 2003). OspA and OspC are important outer surface proteins that are also encoded on plasmids. These proteins play essentials roles in the transmission of the spirochaetes from tick to host (Kurtenbach *et al.*, 2002c). OspA plays a role in the dissemination of the spirochaetes from the midgut of the tick during its blood meal (Yang *et al.*, 2004). While, OspC is believed to be required for

the colonization of the tick salivary glands (Pal *et al.*, 2004b) or early during infection of vertebrate hosts (Fingerle *et al.*, 2000, Schwan, 2003, Schwan and Hinnebusch, 1998). OspA was also the target of the only commercial vaccine that has been available for humans however, it was removed from the market in 2002 due to safety concerns (Abbott, 2006). While there are many essential genes located on the plasmids such as those discussed above, between species there is much variation in number, size and content of the plasmids (Glockner *et al.*, 2006). There are also large amounts of DNA rearrangement, insertions, deletions, and a high number of pseudogenes found on the plasmids (Casjens *et al.*, 2000). Furthermore, it is known that plasmids are often lost during prolonged spirochaete culturing (Schwan *et al.*, 1988). Most of the genes found on the plasmids appear to be unique to the LB species group with less than 10% of plasmid located genes having a predicted function in *B. burgdorferi* (Casjens *et al.*, 2000). This is in contrast to chromosomal genes where the majority of genes are homologous to genes of known function (Fraser *et al.*, 1997).

Table 1.2 Comparison of the genomes of *B. burgdorferi* B31, *B. garinii* PBi (*B. bavariensis*) and *B. afzelii* Pko taken from Kurtenbach and colleagues (Kurtenbach *et al.*, 2010)

Genetic material	***B. burgdorferi*** **B31**	***B. afzelii*** **Pko**	***B. garinii*** **PBi**
Chromosome (kpb)	910	907	905
No. of open reading frames	853	856	832
Plasmid (kpb)	610	507	372
No. plasmids	21	15	11
Linear	12	6	8
Circular	9	9	3

1.6 Identification and typing of LB group spirochaetes

Unambiguous typing systems are essential to understand the ecology and epidemiology of microbial populations. Molecular typing techniques can be categorized as phenotypic or genetic. Phenotypic methods include methods such as serotyping while genetic methods can be further divided into those which use DNA sequence data and those which do not.

1.6.1 Serotyping

Serotyping is the most common phenotypic method used in *Borrelia* and systems based on OspA or OspC are the most well established (Wang *et al.*, 1999b). This technique is based on the differential reactivities of monoclonal antibodies specific to the outer surface protein in question. Using OspA, there are eight serotypes of European strains and species

and a further 11 serotypes of Japanese strains and species. The majority of species are represented by only a single serotype but *B. garinii* is represented by 15 serotypes (Wang *et al.*, 1999b, Wilske *et al.*, 1993, Yanagihara and Masuzawa, 1997) (although one has been redefined as the species *B. bavariensis* (Margos *et al.*, 2009)). OspC has been frequently used in the diagnosis of human cases of Lyme disease, it is more diverse than OspA and many more OspC serotypes have been identified (Wilske *et al.*, 1996b). For example, while OspA provides only one serotype for *B. afzelii* and *B. burgdorferi*, multiple corresponding OspC serotypes have been identified for these two species. Serotyping has provided a simple, reliable approach to identify *Borrelia* species in ticks or patients. However, identification can be hampered by the absence of the outer surface protein as expression of these proteins varies during *in vitro* cultivation (Stevenson and Barthold, 1994, Wilske *et al.*, 1993).

1.6.2 Multilocus enzyme electrophoresis

Multilocus enzyme electrophoresis (MLEE) characterizes bacteria by the relative electrophoretic mobility of several cellular enzymes. Each electromorph is assigned an allele number and so each strain will be assigned a set of allele numbers referring to each enzyme. Then the unique set of allele numbers is assigned an electrophoretic type (ET) and these numbers can be used to assess the make up of the population. MLEE was used to analyse the genetic diversity within the LB group of spirochaetes (Boerlin *et al.*, 1992). This method was most useful during the period when DNA sequencing was not widely available and was useful in measuring overall genetic relatedness and diversity of LB species. It was used to suggest that *Borrelia* are highly clonal (Boerlin *et al.*, 1992). However, the method is labour-intensive and the results are difficult to compare between different laboratories (Wang *et al.*, 1999b).

1.6.3 Genotyping Methods

Methods based on genetic characteristics became very popular due to the fact they could provide more precise information for population studies. Several different methods have been utilized over the last two decades. Most notably, DNA-DNA hybridization has served as the standard method of bacterial taxonomy for many years. In bacterial systematics the accepted phylogenetic species definition was that a species would include strains with greater than 70 % homology with a ΔTm of 5°C or less when tested by DNA-DNA hybridization. Below this value of 70 % homology strains were considered different species (Wayne *et al.*, 1987). Several LB group species have been defined using this

method, these include *B. burgdorferi* B31, *B. afzelii* VS461, *B. garinii* 20047, *B. japonica* HO14, *B. valaisiana* VS116 and *B. lusitaniae* PotiB2 (Baranton *et al.*, 1992, Johnsen *et al.*, 1984, Kawabata *et al.*, 1993). It was shown that in the *Borrelia* genus DNA reassociation was as low as 30 % between the LB group spirochaetes and the relapsing fever spirochaetes (Barbour and Hayes, 1986). Within the LB group spirochaetes DNA reassociation ranged between 48 and 70 % (Baranton *et al.*, 1992, Johnsen *et al.*, 1984, Kawabata *et al.*, 1993). DNA-DNA hybridization has been for many years the standard for bacterial species delineation. However the method requires a specialized laboratory and the number of laboratories that can perform this analysis worldwide is limited. There are also questions about the interpretation and reproducibility of the method (Stackebrandt and Ebers, 2006). Multilocus sequence analysis (MLSA) was put forward as an alternative to DNA-DNA hybridization and this technique has since been used to define several LB group species and this is discussed further in Chapter 4 (Gevers *et al.*, 2005).

Most population and epidemiological studies have been completed using single loci. Before DNA sequencing was widely available several different methods were used for taxonomic purposes as well as to characterize strains within species. For example, randomly amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) were both commonly used for these purposes (for a review see Wang *et al.*, 1999b). The former uses an arbitrary primer set and low stringency PCR conditions to create strain specific arrays of anonymous DNA fragments. RFLP requires PCR amplification of a specific gene such as the rRNA genes or their intergenic spacer regions. The product is then digested using restriction enzymes and the digest is then run on an electrophoresis gel. Both techniques have been used in epidemiological studies (Liebisch *et al.*, 1998, Marconi and Garon, 1992, Mathiesen *et al.*, 1997, Postic *et al.*, 1994, Wang *et al.*, 1998). Both tests are simple, rapid and can be completed without culturing the spirochaetes which means that large scale screening of infected ticks and patients is possible thus facilitating epidemiological studies (Liveris *et al.*, 1996, Rijpkema *et al.*, 1995, Wang *et al.*, 1998). *Borrelia* are problematic to culture not only because the process is unreliable but also culturing can take several weeks to complete as the spirochaetes have a long generation time (Wilske and Schriefer, 2003). RAPD has been used to identify pathogenic subgroups of *B. garinii* found in patients (Wang *et al.*, 1998). It has been demonstrated that RFLP, using the *MseI* digestion, are in accordance with DNA-DNA hybridization suggesting that it can be used as a rapid method that will generate reproducible results (Postic *et al.*, 1994).

Reverse line blot has also been a key method in epidemiological studies of LB species due to it being both, a rapid and reliable method. It uses PCR products of the 5S-23S intergenic spacer region (*rrf-rrl*) for hybridization to membrane bound oligonucleotides that are specific for different LB group species. This method was first used to identify the prevalence of different LB group species in Dutch ticks (Rijpkema *et al.*, 1995). Reverse line blot has greater sensitivity than some other methods for detecting mixed infections and partly for this reason it was a key method in identifying the patterns of host specialization. Kurtenbach and colleagues (2001) observed that there were many mixed infections occurring between *B. garinii* and *B. valaisiana* but mixed infections between either of these two bird related species and the rodent related species, *B. afzelii*, were rare. Later studies by Hanincova and colleagues (2003a&b) confirming the association of *B. afzelii* with rodent species and *B. garinii* and *B. valaisiana* with birds also used this reverse line blot method.

1.6.4 Sequence Analysis

With DNA sequencing becoming a faster and cheaper option, sequences of single gene loci became popular for population and epidemiological studies of the LB group of spirochaetes. Many different genes have been targeted in studies depending on the level of variation required and which species were being investigated. *flaB* encodes a flagellin protein (FlaB) and is located on the linear chromosome. It is well conserved and so has been popular for evolutionary studies and species identifications. Due to its conserved nature, *flab* is present in relapsing fever spirochaetes which can then be used as an outgroup to root phylogenetic inferences. This locus has been used to create an early and reasonably complete evolutionary tree of the LB group of spirochaetes (Masuzawa *et al.*, 2004).

Outer surface proteins are more variable and have, for this reason, often been used in population studies within species. *ospA*, located on a 49- to 70- kb linear plasmid, revealed differences in the levels of homogeneity of LB species (Bergstrom *et al.*, 1989, Vitorino *et al.*, 2008). It was observed that there is great variation in *ospA* in *B. garinii* while there is limited variation in some other species such as *B. afzelii* which is consistent with serotyping studies (Wilske *et al.*, 1996b). Studies have suggested that there is a high degree of linkage between *ospA* and another plasmid gene, *ospC*. These genes occur on different plasmids and so it was suggested that there is very limited horizontal gene transfer in the LB spirochaetes and that they are highly clonal (Dykhuizen *et al.*, 1993,

Wilske *et al.*, 1995). However, analysis of *ospA* has also revealed rare examples of horizontal gene transfer between LB group species (Rosa *et al.*, 1992, Wang *et al.*, 2000).

ospC is located on a 26-kb circular plasmid (Sadziene *et al.*, 1993) and is very heterogeneous with a high degree of molecular polymorphism between species (Jauris-Heipke *et al.*, 1995, Theisen *et al.*, 1993). This locus is rarely used for species identification because, while there may be species specific motifs (Fukunaga *et al.*, 1996a, Jauris-Heipke *et al.*, 1995), the high degree of variation means that many strains appear as inconclusive and in phylogenies strains of the same species do not always cluster monophyletically (Kurtenbach *et al.*, 2002c, Lin *et al.*, 2002, Margos *et al.*, 2009). However, due to the high level of variation, *ospC* has been frequently used in population studies within species, most notably within *B. burgdorferi* (Hanincova *et al.*, 2008, Marti Ras *et al.*, 1997, Qiu *et al.*, 2002, Swanson and Norris, 2008). Wand and colleagues (1999) investigated the diversity of *ospC* and found that allele variation within a local population is almost as great as the variation of the entire species. They proposed that the maintenance of *ospC* variation suggests that this is due to frequency-dependent selection on *ospC* by the host immune system (Wang *et al.*, 1999c).

The region encoding the ribosomal RNA proteins (rRNA) has frequently been used in studies of LB spriochaetes where different regions have been used for various purposes and species. The gene cluster contains a single copy of the 16S rRNA (*rrs*) small subunit followed by tandem repeats of the 23S (*rrl*) and 5S (*rrf)* large subunits as illustrated in Figure 1.4 (Schwartz *et al.*, 1992). There are intergenic spacers (IGS) between the 16S and 23S (*rrs-rrl*) and between the 5S and 23S (*rrf-rrl*) of the repeated pairs. The size of the 16S-23S intergenic spacers varies between species but the 5S-23S spacer is approximately 200 bp long across all species (Ojaimi *et al.*, 1994). This organization of rRNA genes appears to be unique to the LB group spirochaetes (Schwartz *et al.*, 1992) and the non-encoding regions have been widely used for inter- and intraspecies studies of the LB group spirochaetes (Bunikis *et al.*, 2004, Girard *et al.*, 2009, Marconi *et al.*, 1995, Ogden *et al.*, 2008, Postic *et al.*, 1994). The 16S subunit has been used in evolutionary and speciation studies (Fukunaga *et al.*, 1996b, Le Fleche *et al.*, 1997). The 5S-23S spacer is possibly the most common sequence based method of species identification in Europe (Comstedt *et al.*, 2009, Pecchioli *et al.*, 2007, Postic *et al.*, 1994). As a non-coding region it is believed to have limited selection pressures acting upon it and be selectively neutral. It has also been commonly used in a variety of population studies, for example, recently it has been used to identify differences between *B. garinii* strains found in sea birds and those found in passerine birds (Comstedt *et al.*, 2009).

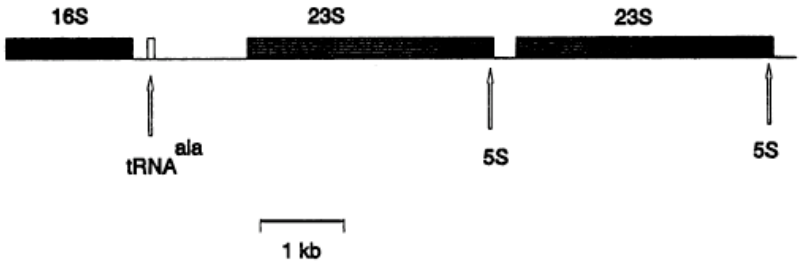


Figure 1.4. Arrangement of the rRNA genes found in the LB spirochaetes by Schwartz and colleagues (1992).

The different loci discussed above as well as several others are used depending on the nature of the study and the discriminatory power required. For this reason different loci tend to be preferred in population and epidemiological studies in United States compared to Europe. In the US there is one predominant LB species causing human disease (*B. burgdorferi*) (CDC, 2008). For this reason most US work uses loci that are useful for intraspecies studies, and more specifically, loci that provide good level of polymorphism for *B. burgdorferi* but not necessarily good for other species or for species identification. Thus many US studies have preferred *ospC* and the 16S-23S IGS (Girard *et al.*, 2009, Hanincova *et al.*, 2008, Marti Ras *et al.*, 1997, Ogden *et al.*, 2008, Swanson and Norris, 2008). The 16S-23S spacer region is much longer than the 5S-23S IGS therefore providing more genetic information (Schwartz *et al.*, 1992).

In Europe several species are prevalent and all four major disease-causing species are endemic in tick populations (Gern, 2008). For this reason species definition has been key for epidemiological and ecological studies, and regions that are conserved enough for interspecies population studies were employed. Thus *ospA* and the 23S-5S IGS region have been most commonly used (Fingerle *et al.*, 2004, Fingerle *et al.*, 2008, Pecchioli *et al.*, 2007, Rauter and Hartung, 2005). Both these loci can be used for species identification and the 23S-5S IGS is the more conserved spacer region so it is possible to compare sequence data of different species.

In Asia species identification is of importance and a variety of loci are used including loci favoured in Europe as well as more conserved loci, such as *flaB* and 16S rRNA (Masuzawa, 2004). This is most likely because fewer population genetic studies have been completed and species prevalence is of primary importance over such a large area considering the broad spectrum of species found across the continent.

Since 2006 several multilocus schemes have been developed to better investigate the population biology of the LB spirochaetes. The greater amount of genetic information permits determination of more subtle differences in populations. Many of these schemes

have been used, as an alternative to DNA-DNA hybridisation, to define new species (Chu *et al.*, 2008, Margos *et al.*, 2009, Postic *et al.*, 2007, Richter *et al.*, 2006, Rudenko *et al.*, 2009a, Rudenko *et al.*, 2009b). However, the majority of them deviate from typical multilocus sequence analysis schemes by combining loci from different evolutionary classes such as housekeeping genes and non-coding regions (Urwin and Maiden, 2003). This is problematic for evolutionary studies because many non-coding regions, such as the IGS, are only found within the LB group of spirochaetes and no outgroup is available to root phylogenetic trees. The different loci may also cause problems interpreting results in population studies (Didelot and Falush, 2007, Urwin and Maiden, 2003). The rational for and development of these schemes is discussed further in Chapter 4.

1.7 Aims and objectives

LB is an emerging zoonotic disease with an increasing number of human cases in several areas of the northern hemisphere including Britain (HPA, 2010b). Due to its global dispersal, many different single locus and multilocus methods have been utilised in the past making comparison of different studies and data difficult and often impossible. Uniting research under a single 'all purpose' scheme will aid the global study of the LB group of spirochaetes. The first goal of this thesis was to aid in the development of a traditional mutlilocus sequence analysis (MLSA) scheme that used regions of housekeeping genes, allowing it to be used for evolutionary studies as well as population studies. The second goal was to utilise the scheme to better understand the evolutionary processes acting on populations of different LB group species. This can be broken down into several more specific aims:

1. To investigate the LB species found in ticks in Britain and Continental Europe.
2. To further develop the newly designed MLSA scheme, based on housekeeping genes and developed by Gabriele Margos and Klaus Kurtenbach, so that all loci can be amplified directly from infected ticks.
3. To assess the genetic population structure of different European LB group species and infer demographic patterns such as population expansions.
4. To investigate the evolution of the LB group of spirochaetes.

Chapter 2: Materials and Methods

2.1 Tick And Lyme Borreliosis Samples

2.1.1 Tick Collection

Ticks were collected by dragging a pale coloured blanket over the woodland undergrowth and leaf litter. After approximately 10 m the blanket was turned over and the ticks were then picked-off using forceps and stored in 70 % ethanol. This process is repeated several times at each collection site and distances were estimated by two steps approximating to one metre. The temperature and humidity were recorded based on the Met Office Uk Observation archives from Filton observation centre (http://www.metoffice.gov.uk/education/archive/uk/). For each drag area the herbage cover was estimate and the general plant types present were recorded. The number and stage of ticks per drag were also recorded. Larvae were not collected.

2.1.2 LB Samples

Questing ticks were collected by blanket dragging between March and November of 2006 and 2007 in England. Three woodland sites in the North-East Somerset and three located around the Latvian capital, Riga, were visited repeatedly over this two-year period, with a median of 5.5 visits per site per year in England and 4.5 visits per site per year in Latvia. Latvian ticks were collected between 2002 and 2007 by Dr. A. Bormane and were couriered to Bath in 70 % ethanol via Dr. M. Donaghy. Additional ticks were collected from 2007 to 2009 from a broader geographical range of British sites, including sites from Scotland and Wales with the assistance of Ms. M. James, Prof. S. Randolph and Dr. A. Dobson. Sites were considered as distinct areas of woodland, with the borders defined by areas of non-woodland such as open fields, roads, rivers or domestic residences. All ticks were preserved in microcentrifuge tubes containing 70% ethanol. The coordinates of all collection sites can be found in Appendix 1. For the phylogenetic and population analyses twenty-two positive questing ticks from Germany and forty-six isolates from the France were also included (see Appendix 1). The German ticks were collected in 2008 and 2009 by blanket dragging by Mr. F. Seelig. The French isolates were from the Institute Pasteur collection and were cultured and DNA extracted by Dr M. Cornet. The LB strains were from ticks collected from 5 regions in France; Alsace, Auvergne, Limousin, Lorraine and Normandy and have been described by Margos *et al*. (2009) and found in Appendix 1.

2.2 Statistical Analyses

All basic statistical analyses were carried out using Minitab 15 Statistical Software (Minitab Ltd., Coventry, UK). Where appropriate Anderson-Darling normality tests were carried out to determine the type of statistical test to use. Only sites with sufficiently large sample sizes were compared statistically (i.e. >70 ticks) for X^2 tests and > 35 ticks per sample for Mann-Whitney U tests.

2.3 DNA Extraction

Total genomic DNA was extracted from all collected ticks by alkaline hydrolysis (Guy and Stanek, 1991). Ticks were individually placed in safe-lock micro-centrifuge tubes containing 120 µl for nymphs or 200µl for adults of 1.25 % aqueous ammonia (NH_4OH) (Sigma Chemical Company, St. Louis, MO, USA) and homogenised with a disposable sterile pipette tip. The tick homogenate was incubated on a heat block for 20 minutes at 100^{o}C. Tubes were removed from the block for 2 minutes to alleviate internal tube pressure. The tubes were then opened and replaced onto the heat block at 100^{o}C until approximately half the liquid had evaporated. Samples were stored at -20^{o}C. Negative controls totalling approximately 5-10 % of the extraction batch were also included in the preparation but containing 100 µl of 1.25 % NH_4OH only.

2.4 Amplification of Single Loci Genes of Lyme Borreliosis Spirochaetes

2.4.1 PCR amplification of 5S-23S rRNA intergenic spacer region of Lyme Borreliosis bacteria

All tick and bacterial DNA samples were subjected to nested PCR amplification of 5S-23S intergenic spacer from the Rijpkema et al. (1995) method. Briefly, the mastermix for the first round PCR were prepared as stated in Table 2.1. Sterile distilled water (SDW) was used in all amplifications.

Table 2.1 Reaction mix for *ospA*, 5S-23S IGS and the housekeeping gene amplification for first and second round nested PCRs. PCR mastermix volumes represent the required constituents for one reaction. The reaction mix component refers to HotStarTaq mastermix kit (Qiagen Ltd., Crawley, UK) for all first round PCR amplification reactions and BioMix Red (Bioline, London, UK) for all second round reactions.

Component	**Working Concentration**	**Volume (µl)**						
		ospA		5S-23S IGS		Housekeeping genes		*recG*
		1st round	2nd round	1st round	2nd round	1st round	2nd round	2nd round
Reaction mix	2 x conc.	6.5	12.5	6.5	12.5	6.5	12.5	12.5
Primer 3'	10 pmol/µl	0.8	1.5	0.8	1.5	1.25	2.5	2.5
Primer 5'	10 pmol/µl	0.8	1.5	0.8	1.5	1.25	2.5	2.5
SDW	N/A	2.9	4.5	2.9	4.5	1	2.5	5
$MgCl_2$	50 mM	0	0	0	0	0.5	0	0
Template	N/A	2	5	2	5	2.5	5	2.5
Total		**13**	**25**	**13**	**25**	**13**	**25**	**25**

Aliquots of 11 µl of PCR reaction mix were dispensed into 0.2 ml thin walled PCR tubes with 2 µl of template and were then amplified in a MJ Research PTC-225 Peltier thermocycler (MJ Research Inc., Waltham, MA, USA) by heating for 15 minutes at 95^{o}C then 26 cycles; 94^{o}C for 20 seconds, 52^{o}C for 20 seconds, 72^{o}C for 45 seconds. After 26 cycles the temperature was held at 72^{o}C for a further 5 minutes final extension then cooled to 15^{o}C.

The second round PCR reaction mix was set up as described in Table 2.1. Aliquots of 20 µl of mastermix were dispensed into 0.2 ml thin walled PCR tubes with 5 µl of template and were then amplified in a PCR thermocycler by heating for 2 minutes at 95^{o}C then 40 cycles; 94^{o}C for 20 seconds, 55^{o}C for 20 seconds, 72^{o}C for 45 seconds. After 40 cycles then a final extension step at 72^{o}C for a further 5 minutes was included. Samples were kept at 15^{o}C. Samples were then loaded and run on an electrophoresis gel as described in section 2.4.

2.4.2 PCR amplification of *ospA* of Lyme borreliosis spirochaetes

Tick and bacterial samples found positive for the intergenic spacer region were subjected to *ospA* nested PCR. Two different *ospA* PCR protocols were used to amplify fragments of *ospA*, initially a fragment of *ospA* was amplified by the nested PCR protocol described by Guy & Stanek (1991). The inner and outer *ospA* primer sets are shown in Appendix 2. Mastermixes for the first round were prepared as stated in Table 2.1. Aliquots of 11 µl of mastermix were dispensed into 0.2 ml thin walled PCR tubes with 2 µl of template and were then amplified in a PCR thermocycler by heating for 15 minutes at 95^{o}C, then 26

cycles of the following; 94°C for 30 seconds, 45°C for 30 seconds, 72°C for 30. After 26 cycles the temperature was held at 72°C for a further 5 minutes then cooled to 15°C.

The second round PCR mastermix was set up as described in Table 2.1. Aliquots of 20 µl of mastermix were dispensed into 0.2 ml thin walled PCR tubes with 5 µl of template and were then amplified in a PCR thermocycler by heating for 2 minutes at 95°C then 40 cycles of the following; 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 60 seconds. After 40 cycles the temperature was held at 72°C for a further 5 minutes then cooled to 15°C. Samples were then ready for storage or to be loaded and run on an electrophoresis gel as described in section 2.4. In later studies a novel set of primers designed by Lencakova and colleagues (2006) were used as they were able to amplify a larger fragment of *ospA* as well as being more reliable at amplifying *ospA* (Lencakova *et al.*, 2006). All PCR conditions remained the same and the new *ospA* primer sequences can be found in Appendix 2.

2.5 Agarose Gel Electrophoresis

1.5 % agarose gels were prepared by adding 100 ml of Tris-acetate-Ethylenediamine-tetraacetic acid (TAE, Sigma) to 1.5 g of agarose (Invitrogen, Paisley, UK). This suspension was then heated in the microwave until all the agarose has dissolved (approximately 120 seconds). When the mixture had cooled, 5 µl of ethidium bromide solution (10 mg/ml, Promega Corporation, Madison, WI, USA) was mixed in by swirling gently and then gel was poured into a gel tray. 28-tooth combs were fitted into the gel. When the gel had set the combs were removed and gel was submerged in a gel tank with 1 x TAE buffer. 8 µl of each sample was added into each well. 5 µl of Marker 2, 50 to 2 kb DNA ladder (Fisher Scientific, Loughborough, UK) was added to the first well of each row. The gel was run at 100 volts, 400 amps for 40 minutes then viewed and imaged on a Bio-Rad Chemi doc (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK).

2.6 Housekeeping Gene Amplification

2.6.1 Primer Design

The final 8 genes for the MLSA scheme were chosen from 30 housekeeping genes that showed approximate sequence variance between 3 and 7 % in microarray analysis of *B. burgdorferi* isolates. This information was based upon on microarray hybridisations data from Dr I. Schwartz, (New York Medical College, USA, unpublished data) (Terekhova *et al.*, 2006). The 8 genes were selected by Dr. G. Margos based on standard MLST gene selection methods (for a review see Urwin & Maiden, 2003). Primers approximately 22 bp

long separated by 600 to 800 bp were designed to the MLSA gene candidates and tested by Dr G. Margos. Nested primer sets were then designed around these original primer sets to increase product yield and allow for spirochaetes to be amplified directly from tick material. Primers were designed using ClustalW alignement in Mega 3.1 (Kumar, S. et. al., 2004) and Oligo Explorer 1.2 (GeneLink Inc., Hawthorne, NY, US). The gene sequences of three *Borrelia* species available at the time (*B. afzelii* Pko, *B. garinii* PBi and *B. burgdorferi* B31) were used to select conserved gene regions for optimal primer design. Primers were synthesised by Invitrogen, Paisley, UK.

2.6.2 PCR Amplification of housekeeping gene regions

Tick and bacterial samples found positive for LB infection using the intergenic spacer region were subjected to the amplification of housekeeping genes. The nested or hemi nested primers for all 8 housekeeping genes are shown in appendix 2, Table A2.1.

2.6.3 First PCR Amplification

For seven of the eight housekeeping genes (*clpA, clpX, nifS, pepX, pyrG, rplB, uvrA*) the mastermix (Table 2.1) and touchdown PCR conditions for first round and second round were identical and are described below. The eighth housekeeping gene, *recG,* required only 2.5 ul of template DNA added to the second round and PCR conditions differed so these samples were run separately. Aliquots of 8 µl of mastermix were dispensed into 0.2 ml thin walled PCR tubes with 2 µl of template shown in Table 2.1. Gene regions were then amplified on their respective PCR thermocycler programmes.

clpA, clpx, nifS, pepX, pyr, uvrA were amplified using a touchdown PCR programme by first heating for 15 minutes at 95^{o}C. Then 95^{o}C for 30 seconds, 55^{o}C for 30 seconds, 72^{o}C for 60 seconds, this sequence was repeated, reducing the annealing temperature by one degree each time until it reached 49^{o}C. At this stage the following sequence was repeated for 25 cycles; 95^{o}C for 30 seconds, 48^{o}C for 30 seconds, 72^{o}C for 60 seconds. After 25 cycles the temperature was held at 72^{o}C for a further 5 minutes then cooled to 15^{o}C.

recG was primarily amplified using the following PCR programme; first heating to 95^{o}C for 15 minutes, then 26 cycles of the following; 94^{o}C for 30 seconds, 55^{o}C for 30 seconds, 72^{o}C for 60. After 26 cycles the temperature was held at 72^{o}C for a further 5 minutes then cooled to 15^{o}C.

2.6.4 Second Round PCR

Mastermixes for the second round of *recG* amplification were prepared as stated in Table 2.1. Mastermixes for the second round amplification of all other MLSA genes were prepared as stated Table 2.1.

clpA, clpx, nifS, pepX, pyr, rplG and *uvrA* were amplified using the following PCR programme; first heating to 95°C for 2 minutes, then 39 cycles of the following; 94°C for 30 seconds, 49°C for 30 seconds, 72°C for 60. After 39 cycles the temperature was held at 72°C for a further 5 minutes then cooled to 15°C.

recG was primarily amplified using the following PCR programme; first heating to 95°C for 15 minutes, then 35 cycles of the following; 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 60. After 35 cycles the temperature was held at 72°C for a further 5 minutes then cooled to 15°C.

2.7 Sequencing

PCR products for sequencing were primarily out sourced to Qiagen Sequencing Services (Hilden, Germany) or Agencourt Biosciences (Massachusetts, USA) for both DNA purification and sequencing. Single loci PCR products, *ospA* and the intergenic spacer (IGS) were sequenced in the forward direction only. The PCR products of the MLSA genes, *clpA, clpX, nifS, pepX, pyrG, recG, rplB* and *uvrA*, were sequenced in both forward and reverse directions. Sequence traces were then viewed, edited and sorted using Seqman (DNASTAR Inc., Madison, USA). The *ospA* gene and intergenic spacer region sequences were searched against GenBank sequence database using the NCBI BLAST function (http://www.ncbi.nlm.nih.gov/BLAST/) to identify the species of the spirochaetal strain.

2.8 Alignments and Constructing Phylogenies

All sequence alignments were generated using MUSCLE Multiple Sequence Alignment Software (Edgar, 2004). Mega 4.0 was then used to visualise the alignments and ensure the alignment remained in frame. Variable sites, mean p-distance and non-synonymous to synonymous substitutions (dN/dS) of genes was also determined using MEGA 4.0. dN/dS ratio was determined using the modified Nei-Gojobori method and Jukes-Cantor model. Preliminary phylogenetic inferences were made using the neighbour-joining method with bootstrap test using 100 repetitions. The nucleotide substitution model was Kimura 2-parameter model, all other default settings remained constant.

2.9 MLSA

All sequences were compared to the MLST database (http://*Borrelia*.mlst.net). Unique alleles and alleles sets were given new allele and sequence type numbers respectively. As there was no allele over-lap between species, the MLSA data was split into species for analysis. *B. burgdorferi* strains were not included in population studies as there were too few in the dataset for population studies.

2.10 Phylogenetic Inferences

Phylogenetic trees were constructed using three different methods, primarily PhyML 3.0 was used via the ATGC Montpellier bioinformatics platform (Guindon and Gascuel, 2003). The evolutionary model to be used in the phylogenetic analysis was determined using FindModel using PAUP tree building method (Tao *et al.*, 2009). For all phylogenies constructed, the general time reversible (GTR) model with gamma-distributed rate variation across sites was selected except for the IGS region where the Hasegawa, Kishino and Yano (HKY) model was selected. The starting tree was a BIONJ tree and the tree improvement method were subtree pruning and regrafting (SPR) and nearest neighbour interchange (NNI). The branch support values were estimated using approximate likelihood ratios (aLRT) and Shimodaira-Hasegawa-like (SH-like) method. All other settings remained as the default settings.

Bayesian phylogenetic inferences were created using Mr Bayes 3.1.2 (Huelsenbeck & Ronquist, 2001). The alignment was partitioned by codon and the appropriate evolutionary model was selected based on the FindModel analysis. Trees were sampled every 50 iterations for single locus gene trees and every 100 iterations for concatenated gene trees. The stop rule setting was implemented and set so that the Markov chain Monte Carlo simulation automatically stopped when the standard deviation of split frequencies was below 0.0095. Burnin values were set to approximately 10 % of the number of trees sampled. Trees inferred using neighbour-joining method were constructed in Mega 4.0. They were constructed with 100 bootstrap replicates and using the maximum composite likelihood model. All other settings remained as default. Trees were viewed using Mega 4.0 and Dendroscope (Huson *et al.*, 2007).

2.11 Data analysis

2.11.1 goeBURST analysis

Allelic profiles, generated by MLST, were separated into species and entered into the comparative goeBURST algorithm (http://goeburst.phyloviz.net/)(Francisco *et al.*, 2009). goeBURST identifies founder STs and their relationship to other STs through single and double locus variants (SLVs and DLVs).

2.11.2 Recombination - Mutation Analyses

The ratio of recombination to mutation events was calculated using two different methods. The first method utilises SLVs to determine whether the single variable locus has changed by a single point mutation or recombination and was previously described by Feil and Spratt, (2001). This method is based on the theory that if a SLV has more than 2 base pair changes at the variant locus, probability would suggest that this is due to recombination while, in the case of 1 base pair change, it is possible it is due to point mutation. Then alleles containing putative point mutations are assessed to see if the allele is present in distantly related strains within the dataset. If it is present then this SLV is also considered to have occurred by recombination as the probability of two identical point mutations occurring in parallel is small.

The second method was calculated using ClonalFrame, a model-based method also designed for the analysis of multilocus data. It infers the likelihood that particular loci have undergone recombination events (Didelot and Falush, 2007). Input files were created for each species using all eight housekeeping genes as described in the ClonalFrame user manual. θ representing the effective population mutation rate, was estimated using DnaSP V5 (Librado and Rozas, 2009) because ClonalFrame is unable to estimate this value. Finally, the concatenated housekeeping genes were analysed using the split decomposion method in SplitsTree3.2 to visualize putative recombination events in the data set (Huson and Bryant, 2006).

2.11.3 Pairwise Mismatch Distribution and F_{ST} Values

The frequency distributions of pairwise sequence mismatches were calculated using ARLEQUIN 3.0 (Excoffier *et al.*, 2005) for demographic and spatial population expansion models. Model fit was calculated using sum of squares deviations. Constant population expansion model was calculated using DnaSP (Librado and Rozas, 2009). The fixation statistic (F_{ST}) estimates the level of differentiation between populations. The output values are defined as being 'low' when between 0.00 and 0.05, 'medium' between 0.05 and 0.25 and 'high' above 0.25 (Freeland, 2005). Pairwise F_{ST} values were calculated for French,

German, British, Latvian and Chinese populations in each of the three species using ARLEQUIN 3.1 and 100 permutations were run to assess the significance of the F_{ST} value.

2.11.4 TREE-PUZZLE Analyses

TREE-PUZZLE V5.2 was used to evaluate tree topologies and the ability of the alignment to resolve the tree topology (Schmidt *et al.*, 2002). Phylogenies were evaluated with respect to the individual and concatenated gene alignments through maximum likelihood tree reconstructions. Quartet sampling and neighbour-joining tree was used for parameter estimations and the GTR model was used for all analyses as this was the model independently selected by FindModel for all alignments (Tao *et al.*, 2009). The base change rate parameters were entered as estimated by FindModel. The model of rate heterogeneity was mixed with 1 invariable and 4 gamma rates. Finally, alpha, the gamma distribution parameter was also entered as estimated by FindModel. All other settings were left as default. Secondly, the ability of the alignments to resolve the tree topology was investigated through estimating the number of taxa quartets that could not be resolved using the alignment data. To do this TREE-PUZZLE quartet puzzling and likelihood mapping were used. For this analysis all settings were changed from default as stated above and the outgroup taxa was set to *Borrelia duttonii*.

2.11.5 PHYLIP

Majority rule consensus trees were created using PHYLIP V3.69 consensus tree programme (Felsenstein, 2005). Trees were equally weighted and were considered as unrooted, all other settings remained as default. PHYLIP TreeDist programme was used to calculate the Symmetric Difference of Robinson and Foulds which show the number of different bipartitions (internal nodes) between two trees (Robinson and Foulds, 1981). Trees were treated as unrooted and all other settings remained as default.

Chapter 3: LB infection prevalence of questing ticks from Britain and Latvia

3.1 Introduction

LB is the most prevalent vector-borne disease in the Northern Hemisphere and there are a growing number of human cases diagnosed in England and Wales each year (HPA, 2010b). The LB group comprises 17 named species that vary in their geographic distribution, host specificity and ability to cause disease in humans (Margos *et al.*, 2009, Rudenko *et al.*, 2009a, Rudenko *et al.*, 2009b). Clinically the different species are of interest as they have been associated with different disease symptoms and other species, such as *B. valaisiana,* are very rarely associated with human disease (Wang *et al.*, 1999b). It is therefore of epidemiological relevance to identify the geographic distribution range of the different LB species.

The LB spirochaetal species also differ in patterns and levels of vertebrate host specialisation. For example, *B. garinii* and *B. valaisiana* are transmitted by avian species while *B. afzelii* is associated with rodents and certain insectivore species (Hanincova *et al.*, 2003a, Hanincova *et al.*, 2003b, Kurtenbach *et al.*, 2002a). *B. burgdorferi* s.s. is a generalist shown to be able to infect both rodent and avian species as well as other hosts (Ginsberg *et al.*, 2005, Hanincova *et al.*, 2006).

In Europe three species are most abundant, *B. afzelii*, *B. garinii* and *B. valaisiana* but these species are not evenly distributed throughout the continent (Kurtenbach *et al.*, 2001). In the Scottish highlands *B. afzelii* is not only present but appears to be the predominant species detected (Ling *et al.*, 2000), whereas in England there are no published data showing the presence of *B. afzelii* in questing ticks, instead, *B. valaisiana* and *B. garinii* are the most abundant species (Kurtenbach *et al.*, 1998). There is, however, a general lack of published data concerning the infections found within tick populations in England and Wales. To the best of my knowledge the last conclusive study investigating LB infections in English questing ticks focused solely on one site in Southern England (Kurtenbach *et al.*, 1998).

The principal vector in Europe is *Ixodes ricinus*, the sheep tick (Burgdorfer, 1984). Abundance and seasonal appearance (phenology) of ticks are a further crucial factors to the ecology of Lyme borreliosis (Kurtenbach *et al.*, 2006). Ixodid ticks feed on a wide range of host species (Estrada-Pena *et al.*, 2005, Talleklint and Jaenson, 1997); immature ticks stages (larvae, nymphs) tend to feed on small animals and adult female ticks feed almost exclusively on deer (Hillyard, 1996). Each fed adult tick may give rise to thousands of

eggs making deer the reproductive host of the tick (Stafford *et al.*, 2003). These conditions lead to the situation that regions absent of deer are also absent of the ticks, hence absence of Lyme borreliosis.

Tick densities, infection prevalence and *Borrelia* species information are all important to understanding the disease risk to the public. Here I show that tick populations are established in the Avon and Wiltshire region and that LB infections are present within the populations. I then screened ticks from a broader range of sites in Great Britain including sites in England, Scotland and Wales and show that infection prevalence varies substantially between sites. I also reveal, for the first time, the presence of *B. afzelii* in England. Finally, I compare my findings in Great Britain to my results from three well-studied sites in Latvia.

3.2 Results

3.2.1 Tick densities at sites in the Bath area

I. ricinus ticks were collected from woodland sites in the Avon and north-west Wiltshire area throughout the spring to autumn period of 2006, 2007 and 2008 as described in section 2.1.1 (Figure 3.1). Temperature and humidity were recorded per collection day and percentage plant coverage of drag area was recorded for each drag as well as the number of ticks collected. Sites were considered as distinct areas of woodland, with the borders defined by areas of non-woodland such as open fields, roads, rivers or domestic residences. See appendix 1, Table A1.2, for table of ticks collected.

The nymphal densities ranged between 0 nymphs per square metre, at the Campus site, and 1.1 nymphs per square metre at Bathampton woods, while the adult density ranged from 0 to 0.21 ticks/m^2. A full list of tick densities can be found in Appendix 1, Table 1.2. Overall there was a median ratio of 9.5:1 nymphs to adults when considering all sites. However, the ratios between sites varied enormously from 14.5:1 nymphs to adults at Bathampton woods, to 0.7:1 nymphs to adults at the campus site. The campus site was the only site where there was a higher median density of adults compared to nymphs.

Figure 3.1. Google map of Europe showing tick collection sites. Labels a to n represent the following collection sites; (a) Inverness, (b) Bathampton, (c) Widcombe Hill, (d) Rainbow Wood, (e) Warleigh Wood, (f) Campus, (g) New Forest, (h) Hazeley Heath, (i) Richmond Park, (j) Exmoor, (k) Rhossili Downs, (l) Kemeri, (m) Babite, (n) Jaunciems.

Due to seasonal differences often observed in tick density data, median nymphal densities from May and September where used to compare different sites (Figure 3.2). Nymphal densities differed between sites but appeared to remain reasonably constant over the different years of collection. Sites tended to have a higher or similar median density in May compared to September. There were no significant correlations of nymph densities with temperature, humidity or herbage coverage at any of the sites (data not shown).

Surprisingly, the tick density data did not follow the expected tick activity curve, which peaks in tick activity in June and July (Randolph *et al.*, 2002), over either year (2006 and 2007). It seems likely that this is due to the unusually wet summer months that occurred during these years, particularly 2007. Only limited amounts of blanket dragging could be carried out in the months of June and July due to rain inhibiting collection. There was also a significant difference in nymphal densities in the second half of the year (July to December) compared to the first half (January to June) of 2007 (Mann-Whitney test, P=0.0145) using data from the three main collection sites (Bathampton, Warleigh, Campus). Bathampton woods was the only site with a sufficiently large sample set in both the first and second halves of 2007 to compare individually and this site also showed a significant difference (Mann-Whitney test, P = 0.0297) between the median nymphal densities of the first half of the year compared to the second.

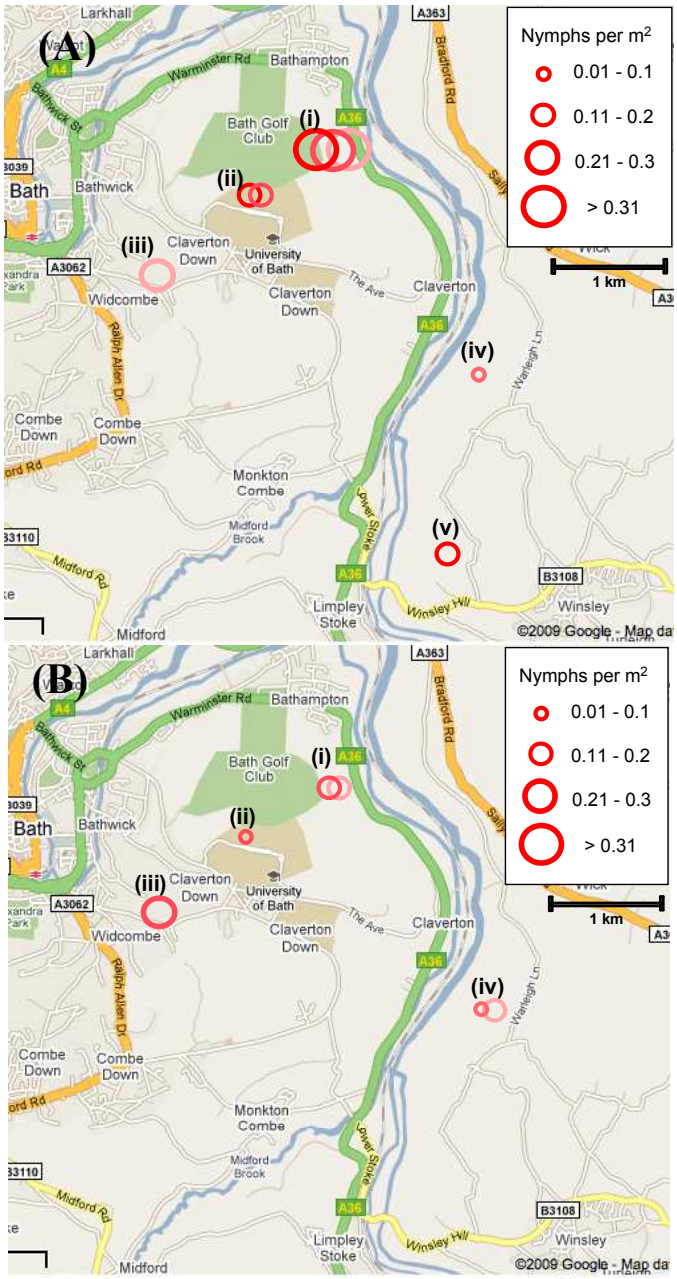


Figure 3.2 Nymphal densities in May and September at collection sites in the Bath area. The map shows mean tick densities recorded in a specific year during the month of (a) May and (b) September. Circles represent the mean nymph density recorded where the larger the circle, the higher the density. The lightest coloured circled refer to data collected in 2006 while the middle red circle are from 2007 and the darkest circles represent data from 2008. Sites shown are (i) Bathampton, (ii) Campus, (iii) Widcombe Hill, (iv) Warleigh and (v) Winsley.

3.2.2.1 *Prevalence of Borrelia infections in field collected* I. ricinus *nymphs*

A total of 1,910 nymphs were collected by blanket dragging at sites in Great Britain in 2006 and 2007, and 5 % of nymphs (n = 96) were positive for *Borrelia* infection. In 2008 and 2009 an additional 827 nymphs were collected from a broader range of sites in Great Britain, and 7.7% (n = 64) nymphs were positive. In Latvia a total of 951 nymphs were collected in 2006 and 2007, and 10.6% (n = 101) were positive. Furthermore, 368 nymphs collected in 2002 in Latvia and screened showed a *Borrelia* infection prevalence of 14.4%. Adult ticks from Britain and Latvia were screened but not included in the prevalence studies, as there were too few to draw meaningful conclusions.

The data were separated by site and year and the infection prevalence for the three main collection sites in England and Latvia is shown in Figure 3.3. Table 3.1 shows the infection prevalence for all major collection sites. While there was a marked difference in

the overall infection prevalence in each country (see above), there was also great variation in the infection prevalence between sites within the same country. There was also a large amount of over-lap in the range of infection prevalence found at sites between the two countries (Figure 3.3). In England, infection prevalence ranged from 0 % infected nymphs at Rhossili downs, Wales, to 12.4 % at Bathampton Woods, Bath and at Latvian sites it ranged from 4% in Jaunciems to 20.8% in Babite (Table 3.1).

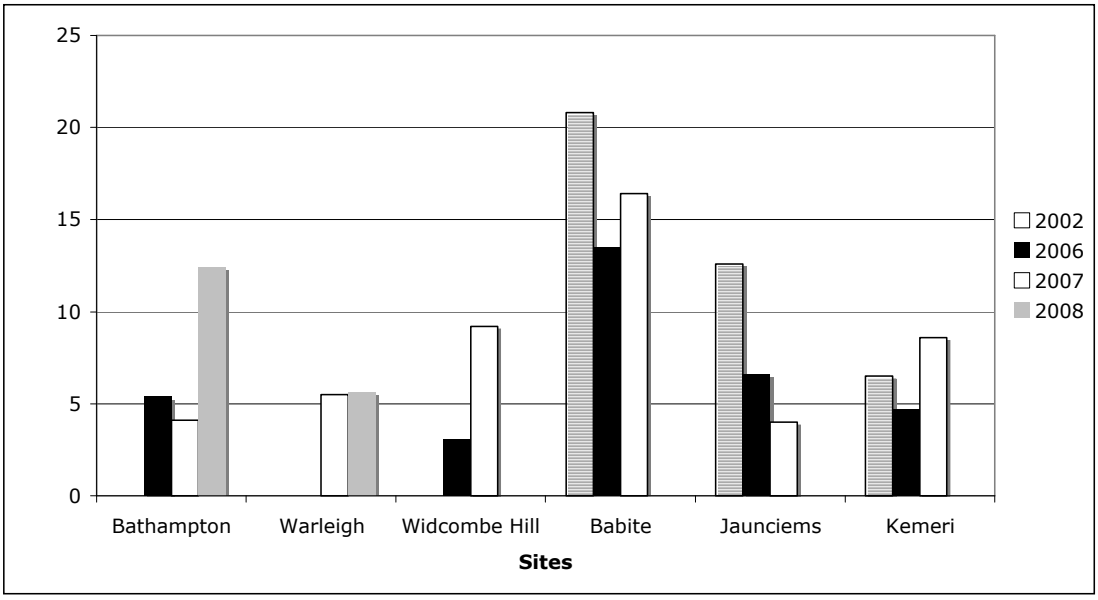


Figure 3.3 Infection prevalence in nymphs as found in different years and sites. Figure shows the 3 main English sites; Bathampton, Warleigh and Widcombe Hill, and the 3 sites in Latvia; Babite, Kemeri and Jaunciems.

Table 3.1. LB prevalence of nymphs collected at sites in Britain and Latvia

Collection Site	Collection year	Nymphs screened	*B. afzelii*	*B. burgdorferi*	*B. garinii*	*B. valaisiana*	*B. lusitaniae*	mixed	Total Positive
ENGLAND									
Bathampton woods	2006	184	0	0	7 (3.8)	2 (1.6)	0	1 (0.5)	10 (5.4)
Bathampton woods	2007	414	0	0	9 (2.2)	6 (1.4)	0	2 (0.5)	17 (5.2)
Bathampton woods	2008	193	0	0	13 (6.7)	5 (2.6)	0	6 (3.1)	24 (12.4)
Eastwood campus	2007	135	0	0	3 (2.2)	1 (0.7)	0	0	4 (3)
Exmoor	2008	175	2 (1.1)	0	1 (0.6)	2 (1.1)	0	0	5 (2.9)
New Forest	2008	134	1 (0.7)	0	1 (0.7)	3 (2.2)	0	1 (0.7)	6 (4.5)
Rainbow woods	2008	82	2 (2.4)	0	3 (3.7)	1 (0.7)	0	3 (3.7)	9 (11)
Richmond Park	2008	71	0	0	1 (1.4)	0	0	0	1 (1.4)
Warleigh woods	2007	110	0	0	3 (2.7)	2 (1.8)	0	1 (0.9)	6 (5.5)
Warleigh woods	2008	72	1 (1.4)	0	2 (2.8)	1 (1.4)	0	0	4 (5.6)
Widcombe Hill	2006	317	4(1.3)	0	1 (0.3)	3 (0.9)	0	2 (0.6)	10 (4.1)
Widcombe Hill	2007	336	13 (3.9)	0	9 (2.7)	6 (1.8)	0	3 (0.9)	31 (6.4)
SCOTLAND									
Inverness	2007	155	3 (1.9)	0	0	0	0	0	3 (1.9)
WALES									
Rhossilli Downs	2009	82	0	0	0	0	0	0	0
LATVIA									
Babite	2002	149	8 (5.4)	2 (1.3)	11 (7.4)	3 (2)	0	7 (4.7)	31 (20.8)
Babite	2006	259	2 (0.8)	0	14 (5.4)	12 (4.6)	0	7 (2.7)	35 (13.5)
Babite	2007	220	0	3 (1.4)	13 (5.9)	10 (4.5)	1 (0.4)	9 (4.1)	36 (16.4)
Jaunciems	2002	127	14 (11)	0	0	0	0	2 (1.6)	16
Jaunciems	2006	106	6 (5.7)	0	0	0	0	1 (0.9)	7 (6.6)
Jaunciems	2007	100	4 (4)	0	0	0	0	0	4 (4)
Kemeri	2002	92	3 (3.3)	0	2 (2.2)	1 (1.1)	0	0	6 (6.5)
Kemeri	2006	127	4 (3.1)	0	1 (0.8)	1 (0.8)	0	0	6 (4.7)
Kemeri	2007	139	5 (3.6)	0	3 (2.2)	4 (2.9)	0	0	12 (8.6)

Proportions of infected nymphs from different years at individual sites were compared to ascertain if there is significant variation between the years. The three Latvian sites appeared to maintain more constant infection prevalence between years and using a X^2 test, no significant difference was found between the infection prevalence in 2002, 2006 and 2007 at any of the Latvian sites. However, Jaunciems showed a marked trend of decreasing infection prevalence over the three years. English sites showed more variation in infection prevalence between years (Figure 3.3). For example while the Bathampton infection prevalence remained fairly constant between the years of 2006 and 2007 (X^2 test, P = 0.570), in 2008 it more than tripled from 4.4 %, in 2007, to 17.5 %, in 2008 (X^2 test, P = 0.000).

Finally, no statistically significant difference was found between the proportion of infected ticks in Spring compared to the Summer/Autumn period at any of the English or Latvian sites of any year and no correlation was found between site infection prevalence and tick density (Pearson correlation P = 0.625).

3.2.2.2 *Species Diversity*

Babite, Latvia, was found to be the most species rich site of all sites studied (Table 3.1) with five different LB species found here, these were *B. afzelii, B. burgdorferi*, *B. garinii*, *B. lusitaniae* and *B. valaisiana*. Excluding Richmond Park, where only a single infected nymph was found, the two sites with the lowest diversity were Jaunciems, Latvia, and Inverness, Scotland, where exclusively *B. afzelii* strains were detected. In England and Wales a total of 9 sites were considered to have a sufficiently large number of screened ticks (> 70 nymphs) to compare the species of *Borrelia*. These sites were Bathampton woods, Eastwood campus, Exmoor, New Forest, Rainbow woods, Richmond Park, Warleigh woods, Widcombe Hill and Rhossilli Downs (one infected adult tick was found here). These sites showed some variation in the LB species identified; at 5 of the 9 sites *B. afzelii* was found and at all of these sites *B. garinii* and *B. valaisiana* were also identified. At one of these sites, Warleigh woods, an extremely low level of *B. afzelii* infection was observed as only one tick was positive for *B. afzelii* out of 217 ticks tested from 2006 to 2008, the strain was found in a nymph collected in 2008. At two of the four *B. afzelii* negative sites only a single infected tick was found at each site (see Table 3.1).

The *Borrelia* species at the three Latvian sites were compared by year and it was observed that at two of the three sites (Babite and Jauciems) there was a reduction in the prevalence of *B. afzelii* over period. In the case of Babite it appears that the level of *B. afzelii* infection had fallen below detection limit while, at Kemeri *B. afzelii* infection

prevalence remained fairly constant (Figure 3.4). Data from the same sites from ticks collected in 2000 and published by Etti and colleagues (2003) was included in Figure 3.4 seems to support this finding.

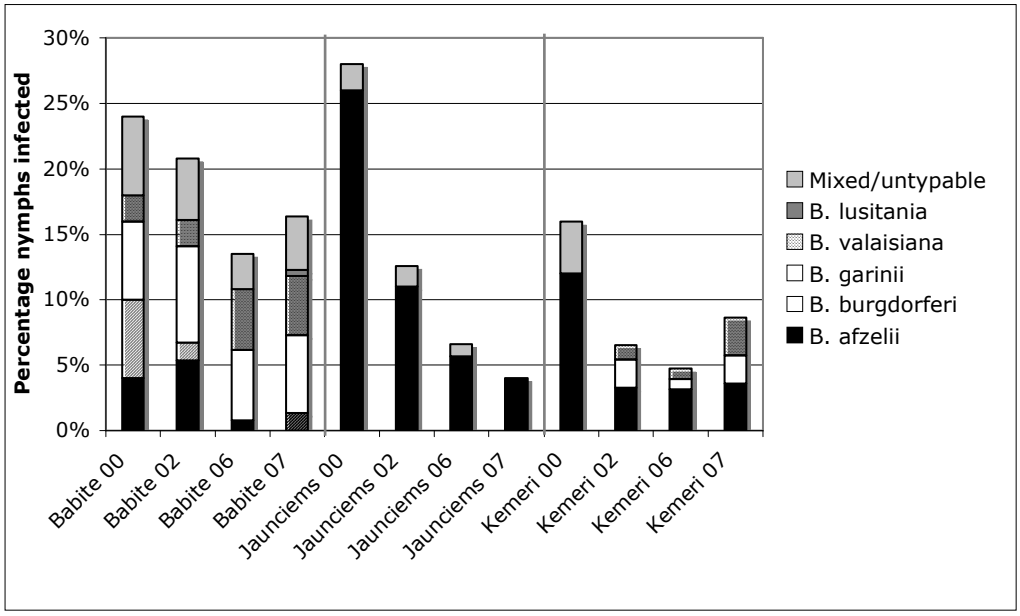


Figure 3.4 The percentage of nymphs infected with the different *Borrelia* species at each Latvian site of each year. For comparison data from Etti and colleagues (2003) who investigated *Borrelia* infections at these sites in 2000 were included in the graph. For simplicity infections that they found untypeable were grouped with mixed infections. For our data we were able to type all strains by species and mixed infections included nymphs that are infected with different strains of the same species as well as different species.

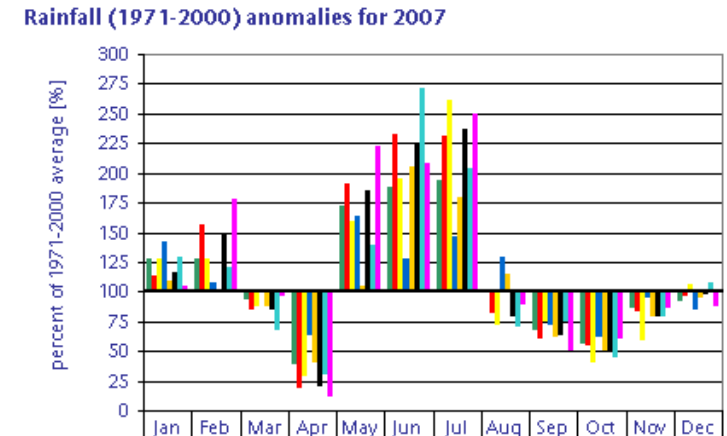
3.4 Discussion

3.4.1 Density of ticks in England

I. ricinus ticks were found at all sites where blanket dragging was attempted in the Avon and Wiltshire area, suggesting populations of this tick species are well established in the region. Sheaves and Brown (Sheaves and Brown, 1995) found questing ticks in the Quantock Hills in South Somerset, a neighbouring county but to the best of our knowledge this is the first published data showing the presence of ticks in the Avon and Wiltshire area (Pietzsch *et al.*, 2005). It has been suggested (Scharlemann *et al.*, 2008) that *I. ricinus* ticks have been emerging in the region over the past 30 years due to the population growth of deer species (Ward, 2005). Ward reveals that deer populations in the UK have been undergoing spatial expansions, including in the Avon and Wiltshire region and as they are reproductive hosts of the tick, this may indicate the tick populations are a fairly recent introduction to the region. However, the lack of published tick collection in the past in the

Avon and Wiltshire region, could either be due to a lack of tick populations or the fact that no collections had been carried out.

The tick densities of the English sites fell within the range found at sites in Continental Europe as well as in previous studies investigating other areas of Britain (Knap *et al.*, 2009, Medlock *et al.*, 2008, Randolph *et al.*, 2002). However, while the consensus of published work shows a tick activity curve over the year, with a peak in activity occurring in June or July in Britain (Knap *et al.*, 2009, Medlock *et al.*, 2008, Randolph *et al.*, 2002), no such curve was observed in our data. It seems likely that the unusually high rainfall in June and July 2007, where more than twice the average rainfall in each of these months (MetOffice, 2010), may have contributed to the unusual densities found in 2007. It is also possible that a more intensive schedule of tick collection would have revealed the expected tick density curve in 2006 and 2008. However, it is not necessarily the case that the rain caused a reduction in tick questing behaviour. Studies by Dautel and colleagues (Dautel *et al.*, 2008) have designed a tick arena which allows for questing tick numbers to be counted directly and it has been suggested by them that rain does not reduce the number of questing ticks and that it is likely to be the blanket dragging method that in ineffective in the damp conditions (Kahl, O., pers. comm., 18th Oct 2008).



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
UK	128	127	93	39	172	187	194	102	67	56	86	92
England	114	156	84	19	192	232	231	81	60	54	84	95
Wales	128	128	87	29	160	195	261	72	74	40	59	106
Scotland	142	107	102	63	164	128	146	130	72	62	95	84
N Ireland	109	100	87	41	105	205	179	115	61	50	79	95
Eng and Wal	117	150	85	21	186	225	237	79	63	51	78	98
England N	128	121	68	31	139	271	204	71	75	44	80	107
England S	104	178	96	11	223	208	249	89	51	60	87	88

Figure 3.5 A MetOffice graph (MetOffice, 2010) showing the percentage of average monthly rainfall that fell in the different regions of UK in 2007. Thus the black line at 100 % represents the average rainfall for that month and bars below/above this line represent lower/higher than average rainfall respectively. The average rainfalls are defined as the mean rainfall between 1971 and 2000 calculated by month.

It is interesting to note that while individual sites appeared to differ in median nymphal density in a given month, the site nymphal densities between years tended to

remain reasonably constant (Figure 3.2). This finding was similar to those of other British studies where there is a much greater variation in nymph densities between sites compared to densities between years (Knap *et al.*, 2009, Medlock *et al.*, 2008, Randolph *et al.*, 2002). It is likely that underlying ecological factors are influencing tick density and that these differ more between the sites than climatic factors cause them to differ between years. Ostfeld and colleague (Ostfeld *et al.*, 2006) have shown that host density (chipmunk and mouse) in one year is a predictor of nymph density the following year. Variations in host type and density at the English sites may also explain the large variations found in the ratio of nymphs to adults. Randolph and colleagues (Randolph *et al.*, 2002) suggests that variations in rates of host-finding at different seasonal periods may influence the number of questing ticks and it seems reasonable to suggest that this theory could apply to different tick stages at the sites as the adult tick stage feeds most commonly on deer and larger mammals, while nymphs feed mostly on smaller mammals such as birds and rodents (Hillyard, 1996). Thus the densities of these hosts at different sites may influence the ratio of questing nymph and adult tick stages.

There was no correlation found between tick density and humidity, temperature or vegetative groundcover. It has been known for a substantial period of time that these factors do, in fact, effect tick questing (Lees and Milne, 1951, Milne, 1944). However, Medlock and colleagues (Medlock *et al.*, 2008) have reported that there are many different determinants for tick densities and highlight the importance of microclimates which I did not investigate in my study. It seems likely that my measures of temperature and humidity were not site specific enough to show any of the expected trends.

3.4.2 Infection prevalence of questing ticks from Britain and Latvia

Overall the infection prevalence in Latvia and Britain was comparable with previous studies (Etti *et al.*, 2003, Kurtenbach *et al.*, 1998). However, a large range in infection prevalence was observed at sites within the same country suggesting the ecological factors at the local level may have a significantly larger impact of infection prevalence than the overall ecology or climate of the surrounding countryside.

It appears that factors affecting tick density may differ from those that effect infection prevalence as there is no correlation between the two. While deer are essential to the survival of tick populations (Stafford *et al.*, 2003, Wilson *et al.*, 1985), they are not reservoir competent so do not play a role in the LB infection cycle and there is little trans-ovarial transmission of LB infections which is likely to explain why tick density is not linked to LB infection prevalence. It is also important to note that, as discussed above,

environmental factors play an important role in tick densities, while are unlikely to effect *Borrelia* populations to the same extent.

Considering that various host species are active at different periods of the year, it is slightly surprising that seasonality did not have an impact on infection prevalence. It may be that any bias in infection prevalence is diluted by unsuccessful questing ticks from earlier in the year or that my sample size was too small to see such a change over the season using X^2.

3.4.3 Species diversity

To date there has been a lack of published data concerning the infections found within tick populations in England and Wales. The last conclusive study investigating LB infections in English questing ticks focused solely on one woodland in Southern England and found only *B. garinii* and *B. valaisiana* in questing ticks and in reservoir hosts (Kurtenbach *et al.*, 1998). In Continental Europe field studies have indicated that *B. afzelii* is common and, in many cases, the dominant species (Etti *et al.*, 2003, Rauter and Hartung, 2005, van Overbeek *et al.*, 2008). It has been suggested previously that tick questing behaviour resulting in a lack of nymphal infestation of mice may prevent the existence of *B. afzelii* in England and Ireland (Gray *et al.*, 1999, Randolph and Storey, 1999), but it has also been a consideration that the lack of sites investigated in England and Wales make it difficult to speculate on this. Here we report for the first time the presence of *B. afzelii* in questing ticks from England where just over half the sites were found to be supporting *B. afzelii* strains. The populations of *B. afzelii* seem fairly localised as not all sites in the Bath area where found to be supporting this species. It seems likely that the presence/absence of *B. afzelii* may be dependant on the host make-up of the individual sites. For example Kurtenbach and colleagues (Kurtenbach *et al.*, 1998) highlight that the site they investigated in Southern England was dominated by game birds, which may mean the site is not able to support rodent related species due to the high number of birds present at the site.

There were large differences in the number of LB group species maintained by different sites. For example at the Latvian site, Jaunciems, only *B. afzelii* strains were found and this has been shown previously at these three sites by Etti and colleagues (2003) and they too suggest this is due to differences in the host populations maintained at the different sites. They noted that ground feeding and ground nesting birds appeared to be more abundant at Babite and Kemeri, where bird related *Borrelia* species were identified in ticks. It is interesting to note that there also appears to be some degree of temporal

variation in species diversity and maintenance at the Latvian sites. Two of the sites showed a gradual reduction in the proportion of ticks infected with *B. afzelii* and this finding was upheld by the study by Etti and colleagues of nymphs from 2000 (Etti *et al.*, 2003). This finding suggests a change occurring in the host populations at these sites however the reasons for these changes are unknown and would require further investigations.

Chapter 4: Development of a Novel Multilocus Sequence Analysis Scheme

4.1 Introduction

4.1.1 The *B. burgdorferi* genome

Borrelia species have unique genomes in that they have one linear chromosome of just over 900 kb but also have a number of linear and circular plasmids. For example, in *B. burgdorferi* B31 an extra 610 kb is spread across 21 plasmids (Casjens *et al.*, 2000, Fraser *et al.*, 1997, Glockner *et al.*, 2006). This is the largest number of extra chromosomal elements found in a bacterium (Casjens *et al.*, 2000). Many of these plasmids are found in all natural isolates of a particular Lyme Borreliosis (LB) group species and, as they may also contain essential genes, it has been suggested that they should considered as mini-chromosomes (Barbour, 1993). However, between species there is much variation in number, size and content of the plasmids. Some plasmids have a much higher content of pseudogenes and non-coding DNA than the chromosome, and the biological relevance of these elements is unknown. Most of the genes found on the plasmids appear to be unique to the LB species group with less than 10% of the plasmid genes having a predicted function in *B. burgdorferi* (Casjens *et al.*, 2000). This is in contrast to chromosomal genes where the majority of genes are homologous to genes of known function (Fraser *et al.*, 1997).

4.1.2 Typing methods for Lyme Borreliosis spirochetes

Unambiguous typing systems are key to investigating epidemiological and ecological patterns, and illuminating the evolutionary processes that shape microbial populations. Several methods have been used to identify strains and species of LB spirochaetes such as restriction fragment length polymorphisms (RFLP)(Belfaiza *et al.*, 1993), single-strand conformation polymorphism (SSCP)(Wienecke *et al.*, 1994) and reverse line blot (Rijpkema *et al.*, 1995). However, to infer phylogenetic relationships between the species, nucleotide sequencing is necessary.

As discussed in chapter 1, in Europe two loci have mainly been used to infer phylogenetic relationships; these are *ospA* and the intergenic spacer region (IGS) *rrf*(5S)-*rrl*(23S)(Postic *et al.*, 1994, Wilske *et al.*, 1996a). When inferring relationships between species outer surface proteins have been problematic as these genes are often under unusual and varied selection pressures which may distort phylogenetic inferences (Fitzpatrick and McInerney, 2005). In contrast, phylogenetic trees generated using IGS

sequences tend to show good resolution at the species level, but limited resolution within the species possibly due to the short length of the locus which is approximately 300 bp. Furthermore due to the fact that both of these loci mentioned here are only found in LB group spirochaetal species, there is no outgroup species to root phylogenetic inferences, therefore there are limited assertions that can be made about the evolutionary relationships of the LB group species.

4.1.2 Multilocus sequence typing and analysis schemes

Multilocus sequence typing schemes (MLST) were originally designed to utilise regions of housekeeping genes that evolved at a more moderate speed compared to more rapidly evolving genes such as the IGS region, which is thought to be under no selection pressures, or outer surface protein coding genes, which are thought to be under variable selection pressures (Maiden *et al.*, 1998). While this means that the number of polymorphic sites per housekeeping gene region is reduced, by combining multiple loci the discriminatory power is increased. One central problem when attempting to understand relationships among bacterial species or populations is posed by horizontal gene transfer. If a single locus representing a particular strain has undergone a recombination event with another strain or species, it would mean that this locus is not representative of the "true" evolutionary pathways of that particular strain genome. In other words, the use of a single locus will infer the evolution of this particular locus but not necessarily the evolution of the organism as a whole. MLST schemes aim to overcome this problem by combining several loci (most often seven) that are scattered across the genome/chromosome. Thus, if one region of the genome has undergone recombination only one or two of the seven genes may be affected. This means primarily, if recombination is occurring it is easier to identify (by comparing base pair changes in the loci of closely related strains or the linkage between genes) as discussed in Chapter 2 (Feil *et al.*, 2000)(Didelot and Falush, 2007). Secondly, using MLST can reduce the distortion in analyses caused by combining genes with different evolutionary histories. This is because in MLST schemes each allele of each gene region is given a unique number so that isolates can then be characterised by a multi-integer number called an allelic profile. Thus regardless of whether a particular strain differs from another strain in a single locus by a single base pair (indicative of mutation) or many base pairs (indicative of recombination), in terms of the allelic profile, the strains will only differ by a single integer number.

Traditionally, the internal fragments of housekeeping genes were selected to be approximately 400-500 bp long, kept in-frame and spread throughout the chromosome to

avoid any local bias that may occur in the bacterial genome (Urwin and Maiden, 2003). The chosen housekeeping genes should also be flanked by genes known to have a similar function as there may be linkage between adjacent genes meaning that genes under varied or strong selection pressures may influence the neighbouring genes. Finally, genes should have a similar level of genetic diversity so that each gene provides a similar contribution in analyses.

Once the genes have been selected and the MLST scheme is in place, sequence data, strain information and allelic profiles are compiled in "virtual isolate collections centres" in the form of online databases (Urwin and Maiden, 2003) such as www.mlst.net. The allelic profiles are given a unique number called a sequence type (ST) allowing for easy reference to particular isolates. The original aim of the MLST concept was to enhance clinical diagnosis, epidemiological monitoring, and population studies (Maiden *et al.*, 1998, Urwin and Maiden, 2003) but the MLST concept has since been broadened to include the analysis of closely related species and this approach has been named multilocus sequence analysis (MLSA) (Gevers *et al.*, 2005, Hanage *et al.*, 2006). MLSA was developed with the aim of allowing for rapid and robust hierarchical classification of all prokaryotic species (Gevers *et al.*, 2005) and has been proposed as a solution to the highly time consuming and difficult method of prokaryote species definition by DNA-DNA hybridization as discussed in Chapter 1 (Bishop *et al.*, 2009, Gevers *et al.*, 2005). Recently a website has been developed to allow the species identification of unknown isolates of Streptococcus species (thought to be a taxonomically challenging group) by entering the sequence data of seven gene fragments (Bishop *et al.*, 2009).

In the LB group of spirochaetes there have been four main multilocus schemes developed (excluding the one described here). Schemes by Bunikis and collegues (Bunikis *et al.*, 2004), Qiu and colleagues (Qiu *et al.*, 2004) and Rudenko and colleagues (Rudenko *et al.*, 2009a) have tended to focus on species found in the United States, with Bunikis and Qiu focusing almost entirely on *B. burgdorferi*, while a scheme developed by Richter considered mostly European species. However, none of these schemes adhere to the strict criteria set out by Urwin and Maiden (2003), described above. All of the schemes combine a variety of gene types including slowly evolving housekeeping genes, non-coding regions, or fast evolving plasmid encoded loci. For example Richter and colleagues used seven loci; *rrs, hbb, groEL, recA, fla, ospA* and *rrf-rrl* (IGS) (Richter *et al.*, 2006) which differ in terms of the selective processes acting upon them, the number of variable sites within these loci as well as the DNA sequence category. The genes are mostly located on the main linear chromosome but *ospA* is located on a plasmid. Richter and colleagues describe the

majority of the genes as informational genes (*rrs, hbb, groEL* and *recA*), but they have also included other categories of loci such as a non-coding region (*rrf-rrl*) and the outer surface protein, *ospA*. This may lead to problems when inferring phylogenies as combining sequence data that are heterogeneous (as loci of different functional categories frequently are) can reduce the power of phylogenetic inference algorithms or even produce erroneous phylogenies (Huelsenbeck *et al.*, 1996).

Furthermore, as mentioned above, the use of the IGS region as well as *ospA* means there is no species available to act as an outgroup to root a phylogeny and to allow for evolutionary inferences. However, these schemes have been used to define several new *Borrelia* species (e.g. *B. spielmanii* and *B. californensis*) by comparing the genetic distance of type strains, based on the concatenated MLSA gene sequence, to the corresponding whole DNA-DNA hybridisation genetic distance data (Postic *et al.*, 2007, Richter *et al.*, 2006).

Here I describe the development of a novel, traditional MLSA scheme based on eight chromosomal housekeeping genes for LB group species and compare the phylogeny generated using concatenated housekeeping gene sequences to the phylogenies of the two traditional single loci, *ospA* and IGS (*rrf-rrl*). The creation of a novel MLSA scheme, based on eight housekeeping genes as shown in Figure 4.1, was led by Klaus Kurtenbach and Gabriele Margos (Margos *et al.*, 2008, Margos *et al.*, 2009) where I designed and optimised nested primers for each of the eight housekeeping genes allowing for successful and consistent amplification of LB spirochaetes directly from tick material (Margos *et al.*, 2009). The advantage is that culturing of strains from tick material was not required, reducing labour as well as the bias that may be created by culturing spirochaetal strains (Norris *et al.*, 1995). While mixed infections cannot be identified, in nymphal *I. ricinus* these approximate to 20 % (Kurtenbach *et al.*, 2001) meaning that the vast majority of tick infections are obtainable using this system. I also reveal differences in the diversity of the different species and investigate rates of recombination.

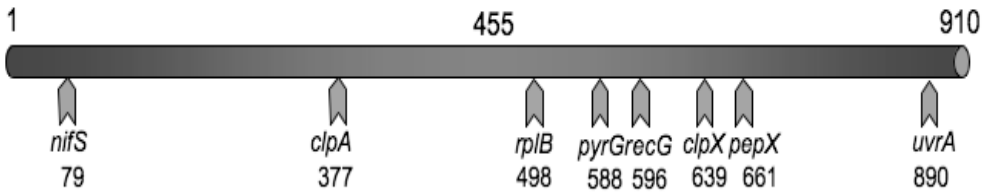


Figure 4.1 Location of eight housekeeping genes of the MLSA scheme on chromosome of *B. burgdorferi.* Numbers refer to kilobase pairs and represent the start location of the genes on the *B. burgdorferi* B31 linear chromosome.

4.2 Results

Questing ticks collected in 2006 and 2007 from England and Latvia were screened for *Borrelia* infections by amplification of the IGS (*rrf-rrl*) region as described in section 2.4.1. The IGS region, *ospA* and housekeeping genes were amplified for all *Borrelia* positive ticks and sequenced as described in sections 2.4 to 2.7, as well as for 46 cultured *Borrelia* strains from France (section 2.4.1). All housekeeping genes were edited using SeqMan before being concatenated. Sequences were aligned using MUSCLE (Edgar, 2004) and Mega 4.0 (Kumar *et al.*, 2004) was used to visualise the alignments, calculate mean pairwise distances (p-distance, π) and non-synonymous to synonymous substitutions (dN/dS) of genes as described in section 2.8. Phylogenetic trees were constructed for all loci as well as for the concatenated housekeeping genes using PhyML 3.0 as described in section 2.10. A single *B. lusitaniae* strain found in a Latvian tick produced poor and short sequence data for the IGS region and *ospA* respectively and so was not included in those phylogenies.

4.2.1 Variation within and among LB species

In general, the mean nucleotide p-distance for *ospA* and IGS were higher than for the housekeeping genes, single or concatenated with one exception, *ospA* for *B. afzelii* (Table 4.1). However, there was a high degree of variation in the mean p-distances of the different species in most loci analysed (Table 4.1). Overall, *ospA* showed the most variation in mean p-distances which was as high as 0.0807 in *B. garinii* and as low as 0.0003 in *B. afzelii* (Table 4.1). The IGS region showed comparatively less variation in mean nucleotide p-distances between the species ranging from 0.0221 in *B. afzelii* to 0.0145 in *B. valaisiana*. It is interesting to note that while *B. afzelii* had the lowest p-distance using *ospA*, it then had the highest p-distance using the IGS region. Individually the housekeeping genes did not show the same pattern of divergence between the species as either the IGS or *ospA* but a similar pattern of divergence was observed between the individual housekeeping genes where *B. afzelii* and *B. valaisiana* had the lowest mean nucleotide p-distances and *B. burgdorferi* and *B. garinii* with the highest mean nucleotide p-distances (with the exception of *pyrG*). The dN/dS values were less than one for all genes, including *ospA* suggesting purifying selection. *ospA* dN/dS ratios were nevertheless substantially higher than those for the concatenated housekeeping genes, except for *B. afzelii*.

Table 4.1 Mean nucleotide p-distance (π) and mean dN/dS ratio of individual genes and concatenated sequence for each LB group species identified except for *B. lusitaniae* and *B. bavariensis* which were excluded since there was only one sample.

Locus	*B. afzelii*		*B. burgdorferi*		*B. garinii*		*B. valaisiana*	
	π	dN/dS	π	dN/dS	π	dN/dS	π	dN/dS
IGS	0.0221	N/A	0.0202	N/A	0.0160	N/A	0.0145	N/A
ospA	0.0003	0.0000	0.0212	0.5942	0.0807	0.7575	0.0551	0.8945
clpA	0.0027	0.0636	0.0113	0.1361	0.0104	0.1651	0.0030	0.3088
clpX	0.0000	0.0000	0.0035	0.0000	0.0082	0.0627	0.0031	0.0148
nifS	0.0010	0.0000	0.0068	0.0194	0.0052	0.1472	0.0017	0.3824
pepX	0.0022	0.0460	0.0088	0.0915	0.0065	0.0929	0.0019	0.5862
pyrG	0.0043	0.0789	0.0037	0.0821	0.0083	0.0721	0.0012	0.0000
recG	0.0021	0.4324	0.0068	0.0558	0.0072	0.2105	0.0020	0.1250
rplB	0.0001	0.0000	0.0045	0.0000	0.0083	0.0239	0.0014	0.0000
uvrA	0.0009	0.0000	0.0084	0.0170	0.0087	0.0557	0.0005	0.0000
Concatenate	0.0017	0.0877	0.0066	0.0536	0.0079	0.0956	0.0019	0.1148

4.2.2 IGS and *ospA* PhyML phylogenetic inferences

The IGS and *ospA* phylogenies shown in Figures 4.2 and 4.3 were not able to group all alleles of the same species on a single branch. In the IGS phylogeny both *B. garinii* and *B. afzelii* appeared as polyphyletic and in the *ospA* phylogeny *B. garinii* was polyphyletic. Furthermore, the IGS phylogeny did not distinguish between *B. garinii* and the newly defined species, *B. bavariensis* (previously known as *B. garinii ospA* serotype 4), which is clustered amongst the *B. garinii* strains. However, for the species which did form monophyletic groups, the branch support values for the species group nodes were all reasonably high (above 75) in both the *ospA* and IGS phylogenies (Figures 4.2 and 4.3). In accordance with low mean nucleotide p-distance, the *ospA* tree also showed a low level of resolution within the species *B. afzelii*. Apart from forming sister clades, the resolution within each cluster was also low for *B. valaisiana* (Figure 4.3).

There was also differences in the tree topologies of *ospA* and IGS. The IGS phylogeny suggested that *B. afzelii* and *B. valaisiana* are adjacent while in the *ospA* phylogeny *B. afzelii* and *B. garinii* are adjacent. However, it is not possibly to infer evolutionary relationships using these genes as it was not possible to define an outgroup for the phylogenies as these genes are not found in any species closely related to the LB group spirochaetes.

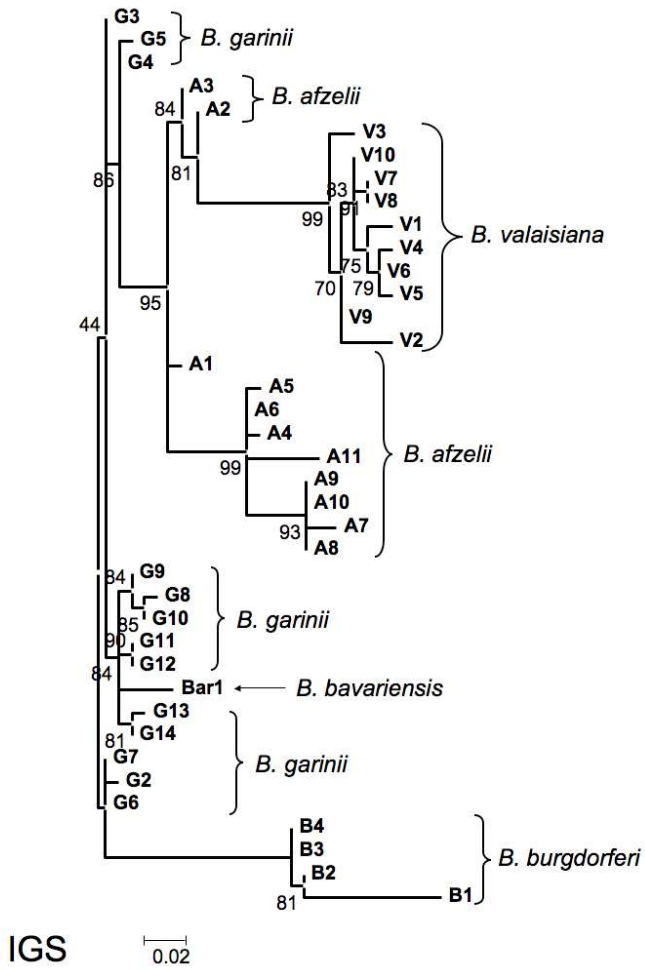


Figure 4.2 PhyML phylogenetic inference of IGS region fragment of *Borrelia* species. The midpoint rooted tree shows all different alleles encountered in the study. Alleles are labelled with letter representing the strain species the allele originates from; *B. afzelii* (A), *B. burgdorferi* (B), *B. garinii* (G), *B. valaisiana* (V), *B. bavariensis* (Bar). Then each allele was arbitrarily assigned a number. A list of strains and which allele number they contain is found in Appendix 3, Table A3.1. The scale bar shows 2% divergence.

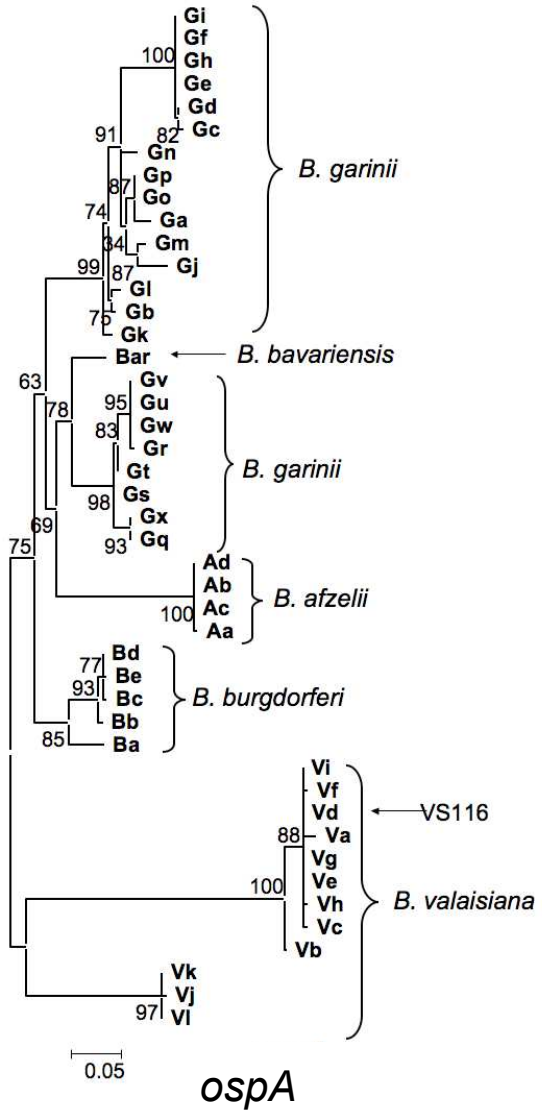


Figure 4.3 PhyML phylogenetic inference of *ospA* gene fragment of *Borrelia* species. The midpoint rooted tree shows all different alleles encountered in the study. Alleles are labelled with letter representing the strain species the allele originates from; *B. afzelii* (A), *B. burgdorferi* (B), *B. garinii* (G), *B. valaisiana* (V), *B. bavariensis* (Bar). Then each allele was arbitrarily assigned a letter (a-z). A list of strains and which allele they contain is found in Appendix 3, Table A3.2. The scale bar indicates 5% divergence.

4.2.2 MLSA Scheme

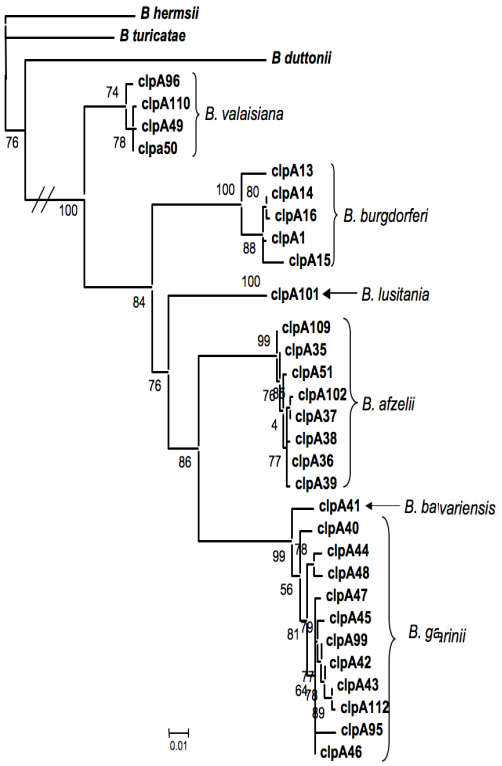
To amplify the housekeeping genes, nested primers were designed to regions which were highly conserved in three species, *B. burgdorferi* strain B31, *B. afzelii* strain Pko and *B. garinii* strain PBi (now renamed *B. bavariensis)* for which genome sequences were available in GenBank. Using the nested primer sets I was able to successfully amplify 96% of strains directly from infected ticks.

4.2.2.1 Investigation of the eight housekeeping genes

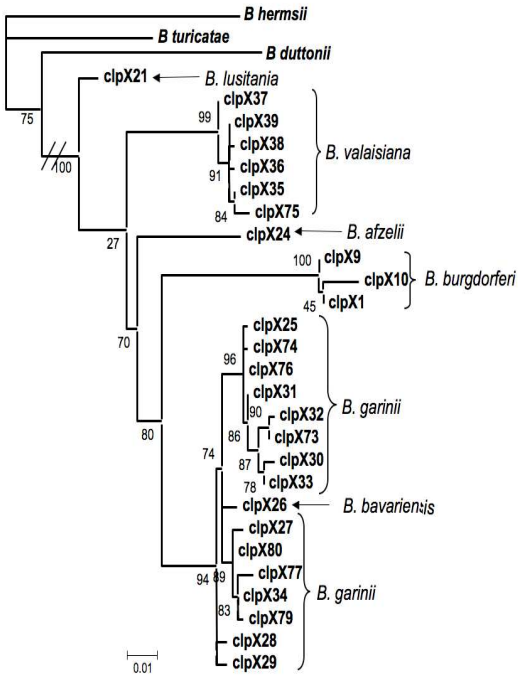
Overall the individual housekeeping genes trees were better able to group alleles from the same species together on a single branch compared to the traditional single loci, *ospA* and IGS. Furthermore, good branch support values for these species groups were achieved for almost all the eight genes (Figure 4.4). Exceptions were the *pyrG* phylogeny (Figure 4.4E), where alleles of the *B. afzelii* species were polyphyletic, and the *pepX* phylogeny (Figure 4.4D), where *B. burgdorferi* alleles were polyphyletic. Six of the eight housekeeping gene

trees were successfully able to distinguish *B. bavariensis* as a separate species and all showed *B. bavariensis* and *B. garinii* as closely related.

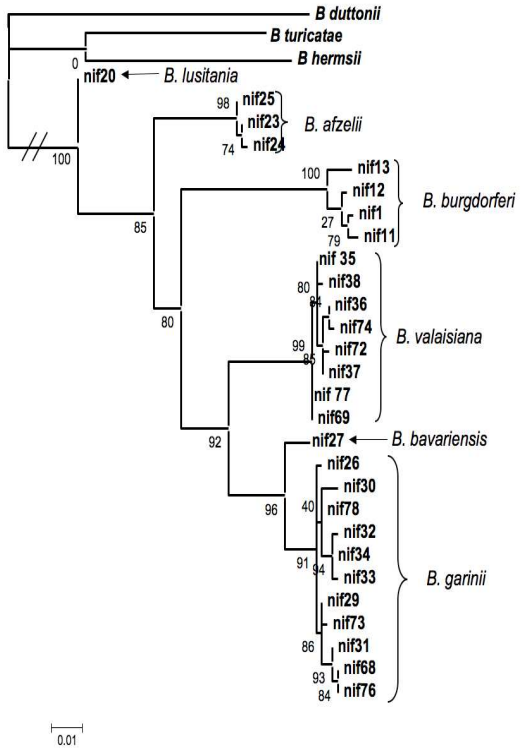
All eight genes trees had different tree topologies when considering the relationships between species groups. The deeper internal nodes that divided the species groups had varying levels of support but were mostly around 70 aLRT. *B. garinii* and *B. bavariensis* were always found to be closely related but their relatedness to other species, as well as among species, was ambiguous. They were found adjacent to all of the other four species at least once in the eight housekeeping gene trees.



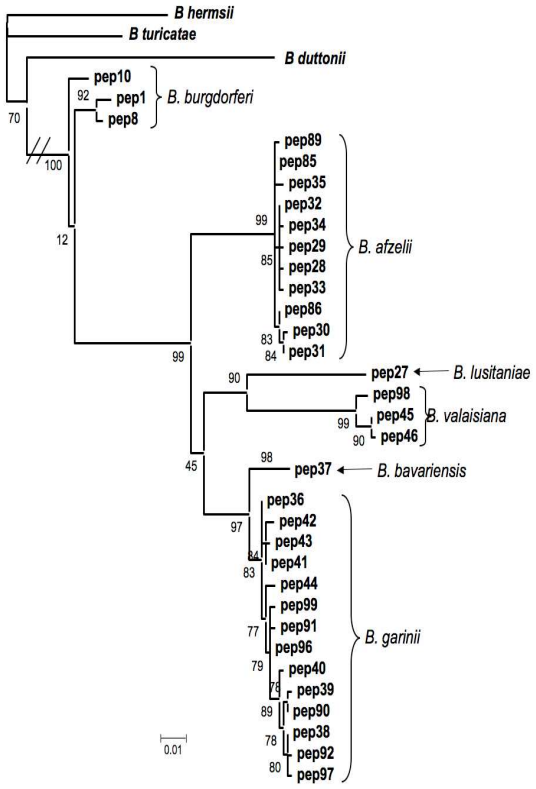
A: *clpA*



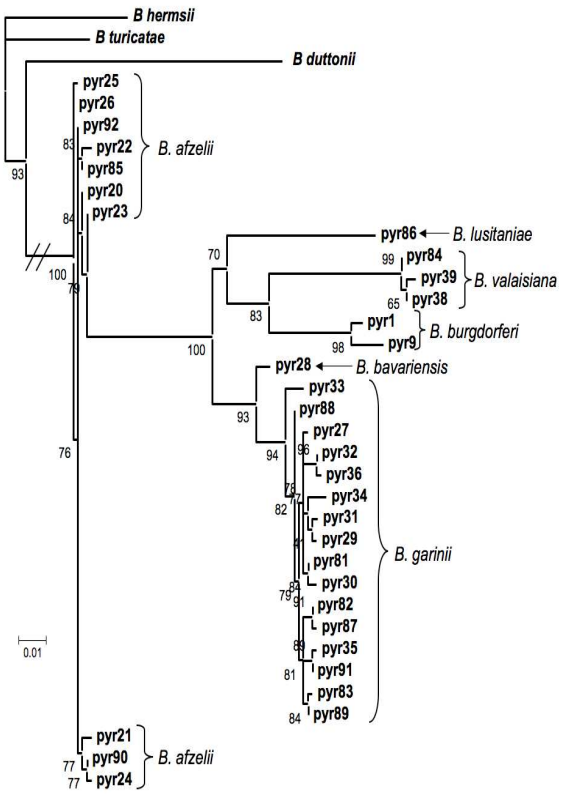
B: *clpX*



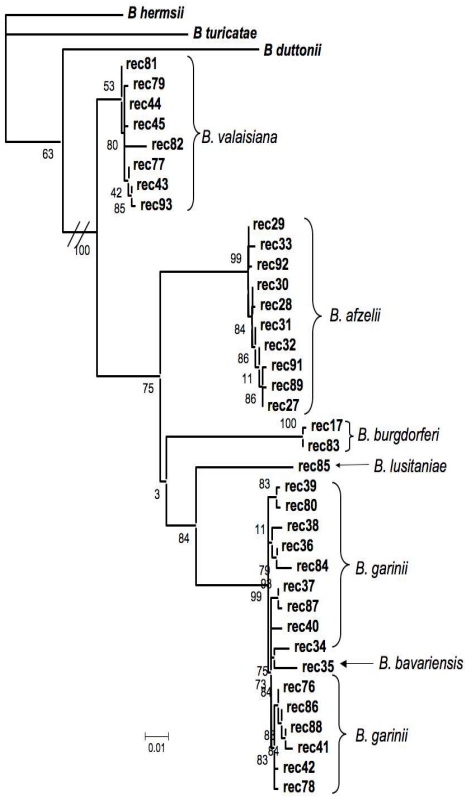
C: *nifS*



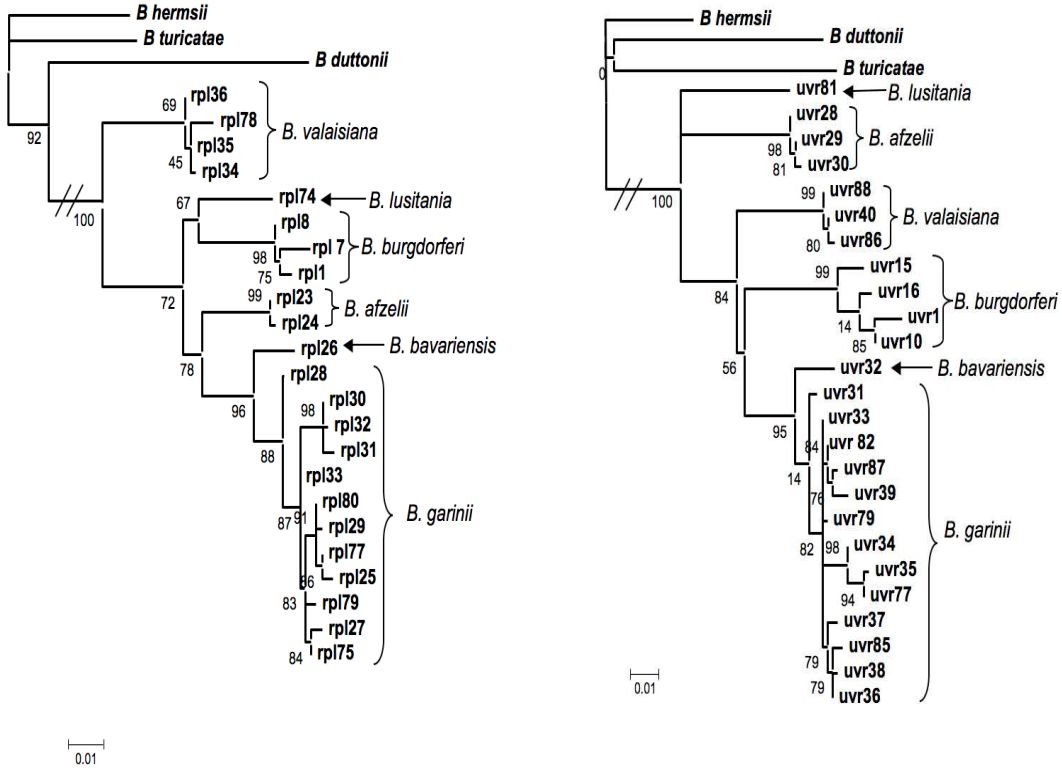
D: *pepX*



E: *pyrG*



F: *recG*



G: *rplB* **H: *uvrA***

Figure 4.4. PhyML phylogenies of individual housekeeping genes; A, *clpA*, B, *clpX*, C, *nifS*, D, *pepX*, E, *pyrG*, F, *recG*, G, *rplB*, H, *uvrA*. Phylogenies show all alleles identified in the strains from this study for each gene. The branch support values are calculated using aLRT method (maximum value of 100). The branches of the outgroup species, *B. hermsii, B. turicatae* and *B. duttonii* are not to scale as indicated by the slashes. Scale bars show 1 % divergence.

4.2.2.2 Concatenated housekeeping gene phylogeny

In the concatenated housekeeping gene tree each species was monophyletic (Figure 4.5) and furthermore, the branch support values for nodes defining a species group have high confidence values at 99 or 100 aLRT. Each species shows significant divergence from the others indicated by the long branches that the species strains cluster at end of and no strain fell between or outside of these well-defined species groups. Furthermore there was no allele overlap between the different species (Appendix 4).

The divergence within the species groups appears to vary greatly where *B. afzelii* and *B. valaisiana* show very little divergence while *B. burgdorferi* and *B. garinii* have much deeper branching (Figure 4.5). This is consistent with and further illustrated by the mean nucleotide p-distance values shown in Table 4.1 where both *B. afzelii* and *B. valaisiana* are less diverse than *B. burgdorferi* and *B. garinii* using the concatenated sequences.

Branch support values for two of the deeper internal nodes were extremely poor (less than 10 aLRT). These low values are most likely due to the fact that the deeper

internal nodes that represent speciation events are on very short branches indicating there are limited informative nucleotide sites representing these deep branches. Furthermore, the disparity in topology between the different single loci (Figure 4.4) will have contributed to the low branch support values in the concatenated gene tree discussed above.

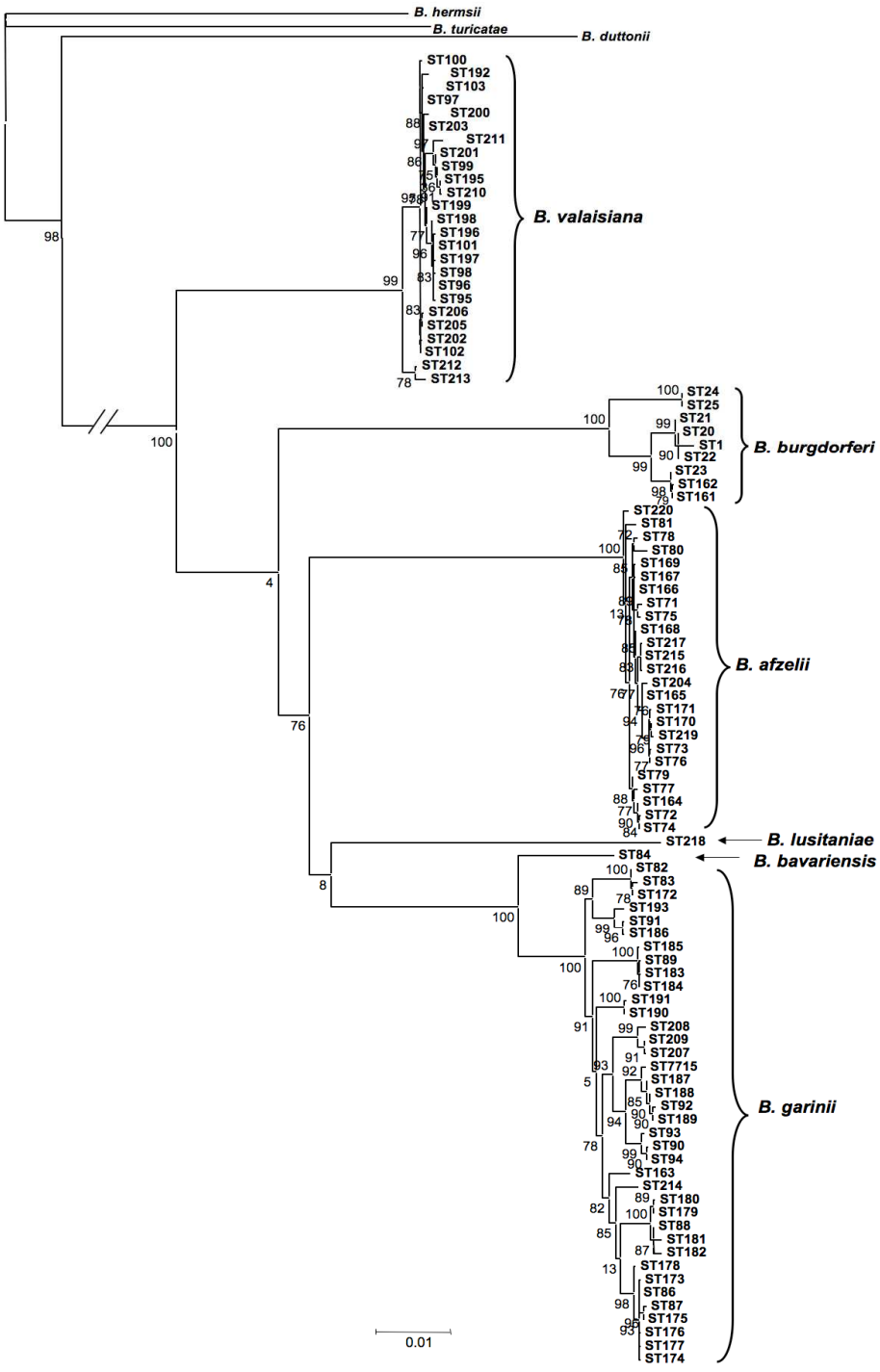


Figure 4.5 PhyML phylogenetic inference the concatenated sequences of the housekeeping genes of STs for all LB group species identified in this study. The branch lengths of the outgroup species; *B. hermsii*, *B. turicatae* and *B. duttonii* are not to scale indicated by the slashes. The branch support values are calculated using aLRT method with a maximum value of 100. The scale bar shows 1% divergence.

4.2.2.3 Recombination

The ratio of putative recombinational to mutational changes (r/m) was calculated for all STs using two different methods. Firstly, r/m was calculated using ClonalFrame and secondly, calculated using Feil per-site method (Feil *et al.*, 1999) both are described in detail in section 2.11.2.

Both tests suggested that mutation was much more common than recombination (Table 4.2). ClonalFrame estimated particularly low r/m values for *B. afzelii* and *B. valaisiana*. In *B. garinii*, however, recombination was estimated to be higher than mutation using ClonalFrame, although with exceptionally large 95 % confidence intervals. Feil per-site method suggested *B. garinii* may have a similar r/m ratio to *B. afzelii*, again suggesting that there are few recombination events compared to mutations in this species, too.

Finally, a network analysis using the software suite SplitsTree was performed using the alleles for all STs identified in this study as described in section 2.11.2 (Figure 4.6). This analysis showed splits in two of the species, i.e. *B. burgdorferi* and *B. garinii*, indicating that recombination may occur in the main chromosome albeit at low level. This supports low r/m estimates as there are very few splits within the tree and suggests that there is a high level of linkage in the main linear chromosomes of LB group species.

Table 4.2. Estimated recombination to mutation rate (r/m) of three LB group species using two different methods of calculation; firstly using clonal frame and secondly using Feil per-site method. Clonal frame also provides a region of 95 % confidence for r/m.

Method		***B. afzelii***	***B. garinii***	***B. valaisiana***
Clonal frame	r/m	0.0531	1.2447	0.0263
	95% credibility region	0.0001 to 0.0436	0.0091 to 7.6637	0.0001 to 0.0768
Feil per-site	r/m	0.429	0.417	0.125

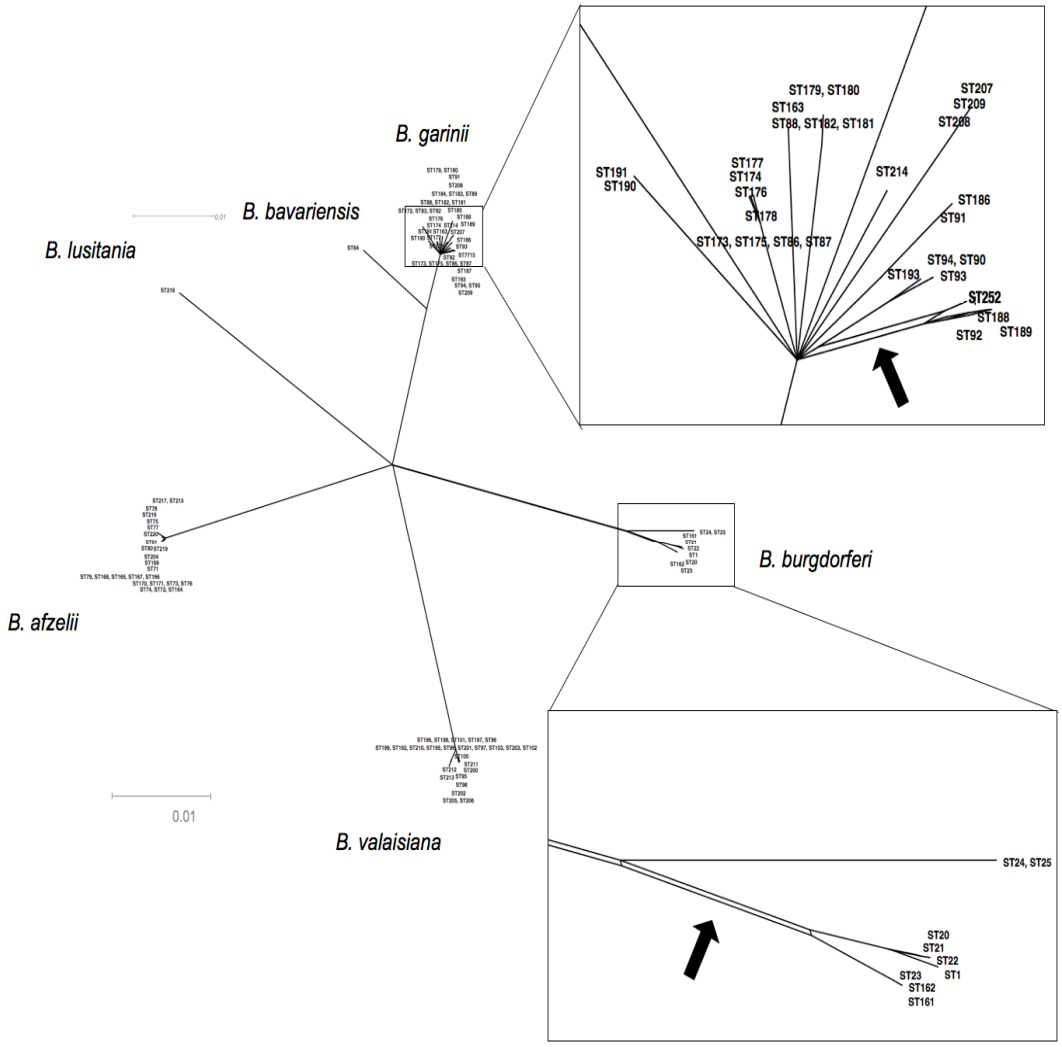


Figure 4.6 SplitsTree phylogeny showing unrooted tree of STs and showing splits in the tree suggesting possible recombination events. All splits in tree have been enlarged and are shown in the boxes. Arrows indicate the locations of the two splits.

4.4 Discussion

In this chapter the successful development of a MLSA scheme for LB group of spirochaetes based on eight housekeeping genes is described. The scheme has been optimised so that *Borrelia* DNA can be amplified directly from infections in ticks, removing any biases created through culturing. I show the results of the scheme using sequence data from infected ticks from Latvia and England as well as from cultured strains from France. Comparison with single locus phylogenies demonstrates the versatility of MLSA in understanding the epidemiology of LB spirochaetes with respect to grouping alleles from the same species in a monophyletic manor, improving resolution within species, identify populations within species and estimating the levels of chromosomal recombination and mutation.

4.4.1 Phylogeny of concatenated housekeeping genes

The MLSA scheme provides the benefit of defining outgroup species as the housekeeping genes are also present in the relapsing fever spirocheates, permitting the use of *B. hermsii, B. dutonii* and *B. turicatae* to root phylogenies. This allows for inferences about the order in which the LB species evolved. However, ascertaining this order proved difficult due to the low branch confidence values in the ST phylogeny (Figure 4.5). Several factors may be responsible for this: i) internal branches representing species divisions are extremely short, representing less than 1 % divergence suggesting that the speciation events, in evolutionary terms, occurred in quick succession. This is further indicated by the star-like phylogeny, indicative of rapid speciation events, in the SplitsTree diagram (Figure 4.6). Thus there are limited mutations existing in the sequence today that represent these intermediary species. ii) These short branches may also be suggestive of incomplete lineage sorting also known as hemiplasy (Avise and Robinson, 2008, Maddison, 1997). This occurs when polymorphisms are maintained in a gene through two or more speciation events thus giving the impression of a different topology (Figure 4.7, (Maddison, 1997). Incomplete lineage sorting may provide an explanation why the single gene trees produced such a wide array of topologies. This problem is then compounded by the fact that each of the LB group species cluster at the end of a relatively long branches potentially erasing changes that did represent the intermediary species. iii) The limited number of genes (8 housekeeping genes) may not be enough to clearly define the topology and furthermore this analysis only included 5 of the 17 known species and the inclusion of more species or genes is known to improve branch confidence (Page and Holmes, 1998).

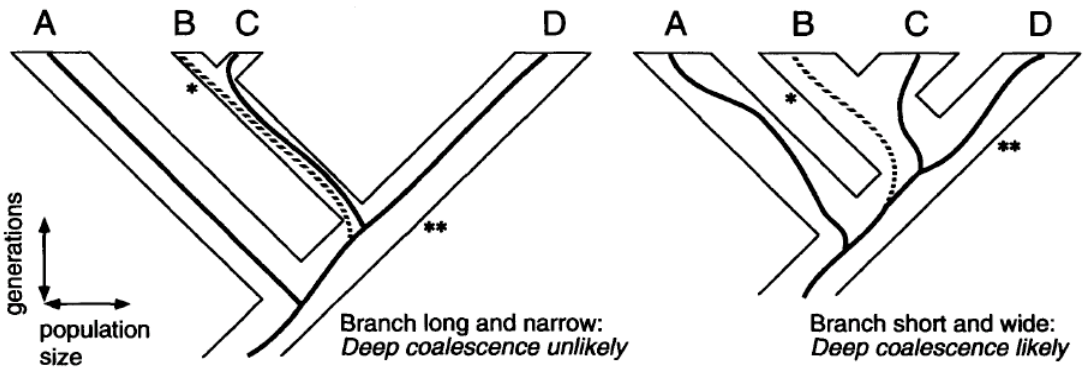


Figure 4.7. Illustration from Maddison (1997) showing hemiplasy. This is represented by the two alleles, B and C, that fail to coalesce until prior to the speciation event at **. Thus a gene tree based on this locus alone would suggest the topology of the dotted and single internal lines while the true tree topology is represented by the rigid margin. The two diagrams together illustrate why this deep coalescence is more likely to occur when branching is short and the ancestral population size is greater (right).

The phylogeny by Margos and colleagues (2009) was extremely similar to the topology to the ST tree presented here. This was expected as the two trees are based on the same MSLA scheme and both included the same cultured French strains but the tree by Margos and colleagues had higher branch support values (Margos et al. 2009). It is possible that the improved branch support values in this phylogeny are due to the additional species groups included in that phylogeny (in addition to the species in my tree, it also includes *B. spielmanii* and *B. bissettii*). Indeed, in a phylogenetic inference presented in chapter 6 the use of additional species does show improved branch support values.

Several other phylogenies have been published of LB group species using other MLSA schemes. Rudenko and colleagues (2009b) and Richter and colleagues (2006) have both published unrooted or midpoint rooted phylogenies as their schemes have included *ospA* and IGS sequence region which are unable to define an outgroup for their trees. Both produced different topologies compared to the concatenated housekeeping gene trees and to each other.

4.4.2 The IGS region and *ospA*

Both *ospA* and IGS have been used in population studies (Grego *et al.*, 2007, Pecchioli *et al.*, 2007). Vitorino and colleagues (2008)(Vitorino *et al.*, 2008) compared various single loci, including IGS and *ospA*, to the housekeeping gene MLSA scheme in their ability to differentiate between *B. lusitaniae* strains from different regions of Portugal. They found that *ospA* produced no clear phylogeographic signal while the IGS marker showed some degree of structuring in the strains but the MLSA scheme was by far the most powerful method for describing the geographic structuring of strains. My study found that *ospA* had differing resolutions within the species of the LB group suggesting that the locus may be more useful for population studies in some LB species than in others. For example *ospA* shows almost no variation among the strains of *B. afzelii* while *B. garinii* show much more variation (Table 4.1). The reason for the highly conserved nature of *ospA* in *B. afzelii* is difficult to interpret. One could speculate that the limited variation suggests the gene, or the plasmid it is located on, may have been recently (in evolutionary terms) acquired and that this gene or plasmid has reached fixation in the European *B. afzelii* population. Several of the chromosomal housekeeping genes also showed very low level of variation as well, such as *clpX* where no variation was observed, suggesting that perhaps the very low variation is genome wide and that the population may have been through a severe bottleneck at some point in the recent evolutionary history.

The IGS region showed more consistent variation between the species which is most likely due to it being a non coding region with few selective pressures acting upon it. Postic and colleagues (1994) first proposed this IGS region for use in epidemiological studies via RFLP, describing it as a simple and useful tool in the study of *Borrelia* diversity. True to this prediction, it has been frequently used in studies of *Borrelia* (Comstedt *et al.*, 2009, Masuzawa *et al.*, 2001, Pecchioli *et al.*, 2007, Rijpkema *et al.*, 1995) however, as Vitorino and colleagues revealed, there is a limit to its ability as a marker in population biology studies most likely due to the short length of the region (Vitorino *et al.*, 2008).

The MLSA scheme has also been utilised for population studies by Hoen and colleagues (2009) who were not only able to show that populations from the Northeast and the upper Midwest of the United States were genetically distinct. These authors also demonstrated an ancient population size expansion using the housekeeping gene scheme suggesting that *B. burgdorferi* was present in the US for several thousand if not millions of years (Hoen *et al.*, 2009). In Chapter 5 I will discuss differences in populations within species with different host specialisations, which was made possible by using the housekeeping gene MLSA scheme.

4.4.3 Species Definition

The tree constructed using the concatenated housekeeping gene sequences grouped species monophyletically. In general multilocus schemes have been superior at defining species groups compared to the single loci (Bishop *et al.*, 2009, Gevers *et al.*, 2005) and have been used to define new species within the LB group complex (Chu *et al.*, 2008, Margos *et al.*, 2009, Postic *et al.*, 2007, Richter *et al.*, 2006, Rudenko *et al.*, 2009a, Rudenko *et al.*, 2009b) suggesting MLSA may in the future become the tool of choice for species definition as it is both more attainable and reproducible than the "gold standard" DNA-DNA hybridisation.

The fact that LB group species form distinct ecotypes that are largely characterised by their host range supports the theory that bacterial strains can form distinct ecotypes which are effectively species. The extant species were at the end of long branches forming distinct clusters and there were no examples of strains that fall between these species groups even though I completed fairly extensive sampling of environmental samples from both Latvia and England. *B. valaisiana* and *B. garinii* are the only two distinct species included in my study that share the same host range and occur sympatrically (Taragel'ova *et al.*, 2008) but are very distinct species and not closely related. This suggests host

specialism to bird species has arisen more than once in LB spirochaetes and it has been suggested that these strains may have evolved allopatrically which may explain their pronounced genetic distance (Margos *et al.*, 2009).

With the sample set tested here, both traditional loci (*ospA* and IGS) were unable to group all alleles from strains of the same species onto a single branch for all species investigated. However, previous studies using the IGS region have shown that this region can form monophyletic groups for strains of the same species (Pecchioli *et al.*, 2007, Richter *et al.*, 2006). This did not occur in my data possibly due to the short fragment length. For example, Richter and colleagues used a fragment that was 197 b.p. long while the largest fragment length that allowed me to include the largest number of strains was 180 b.p.. This may have sufficiently reduced the number of phylogenetically informative sites to make some species appear polyphyletic. In contrast the *ospA* region has, in the past, been shown to be unable to group alleles from the same species one a single branch (Kurtenbach *et al.*, 2002c). Wang and colleagues (2000) found two distinct *B. valaisiana ospA* clades and suggest that they arose due to a recombination event.

The housekeeping loci were selected from regions across the chromosome (Figure 4.1) so that recombination events that occur in different regions of the genome are more likely to be identified in the housekeeping gene data (Feil *et al.*, 1999). The amount of recombination occurring in the LB group spirochaetes has been an area of considerable debate. Early studies suggested that the LB group species are highly clonal and that recombination events are extremely rare (Boerlin *et al.*, 1992, Dykhuizen *et al.*, 1993). While, later studies suggest horizontal gene transfer seems to occur between LB group species but this study was investigating plasmid genes (Wang *et al.*, 1999a). In my analyses I observed no chromosomal gene transfer between species. This was supported by the SplitsTree analysis in Figure 4.6 and the monoplyetic nature of the of the individual gene trees in Figure 4.4. However, the polyphyletic nature of the *ospA* tree in Figure 4.3 may be an indication of rare cases of interspecies recombination in plasmids (Wang *et al.*, 2000).

The intra species recombination analyses completed in this study suggested that *Borrelia* species do have exceptionally low ratio of chromosomal recombination to mutation events in comparison to other bacterial species. The majority of bacterial species, summarised by Vos and Didelot (2009), have higher or relatively equal levels of recombination compared to mutation present in their genomes using ClonalFrame analysis. *B. afzelii* and *B. valaisiana* r/m values are most similar to the lowest species summarised by Vos and Didelot which was also a spirochaetal species (*Leptospira interrogans,* r/m

0.02, (Vos and Didelot, 2009). While the ClonalFrame value for *B. garinii* was unreliable due to the exceptionally large confidence intervals, the Feil method suggested that true r/m maybe very similar to *B. afzelii*. If all three species are considered to have similarly low r/m ratios, this may suggest that the LB group species as a whole may be comparatively clonal compared to many other bacterial species. This may be due to the lifestyle of the LB species group where due to host specialisation, different species rarely coexist in the same environment for any significant period of time. I have shown that the majority of nymphs are infected with a single strain (Chapter 3) suggesting that there are limited strains of the same species infecting the same tick and so there may also be limited opportunity for gene transfer within species as well as between species.

Chapter 5: Populations Biology of LB Group of Spirochaetes

5.1 Introduction

The principal LB vector in Europe is the sheep tick *Ixodes ricinus* (Burgdorfer, 1984), which transmits the LB group spirochaetes between vertebrate hosts, including both avian and mammalian species (Kurtenbach *et al.*, 2002a). LB species tend to vary in terms of host specificity, and many are associated with specific disease symptoms. For example, *B. afzelii* is most frequently linked with skin manifestations (Canica *et al.*, 1993), *B. garinii* with neuroborreliosis (Ornstein *et al.*, 2001, Rijpkema *et al.*, 1997, Ruzic-Sabljic *et al.*, 2001) and *B. burgdorferi* s.s. with arthritic symptoms (Ornstein *et al.*, 2001, van Dam, 2002, van Dam *et al.*, 1993). Other species, such as *B. valaisiana,* are very rarely associated with human disease (Wang *et al.*, 1999b).

Host specialisation is an important factor in vector-borne disease, and different pathogens show varying levels and patterns of host specialisation. For example, western tick-borne encephalitis virus is only transmissible via rodent hosts (Randolph *et al.*, 1999), while West Nile virus transmission depends on birds (Granwehr *et al.*, 2004). A true understanding of the epidemiology of many zoonoses can only be achieved by considering the varied ecological adaptations of the pathogens, particularly differences in host specialisms. In the LB group of spirochaetes, *B. garinii* and *B. valaisiana* are transmitted by avian species while *B. afzelii* is associated with rodents and certain insectivore species (Hanincova *et al.*, 2003a, Hanincova *et al.*, 2003b, Kurtenbach *et al.*, 2002a). *B. burgdorferi* s.s. is a generalist species, known to infect both rodent and avian species, as well as other hosts (Ginsberg *et al.*, 2005, Hanincova *et al.*, 2006). This variation in host specialisation makes the LB group of spirochaetes an ideal model to directly contrast the effects of host specialisation on the migration of pathogens. As ticks do not move over large distances independently (Falco and Fish, 1991), the spread of LB spirochaetes is likely to be linked to the migration of their hosts (Kurtenbach *et al.*, 2002b). Species that are maintained by rodents are therefore predicted to show more limited migration than those associated with birds. In addition to being of public health importance, the delineation and monitoring of the geographic ranges of the different LB species also provides an opportunity to examine in more general terms the role of host ecology in the epidemiology of vectored zoonoses. Here I test the prediction that host migration determines spirochaete biogeography by characterising different LB species from sites in Great Britain, Latvia, Germany and France using Multilocus Sequence Analysis (MLSA).

The distribution of the 17 named species differs greatly across the northern hemisphere. In Europe three species are most abundant, *B. afzelii*, *B. garinii* and *B. valaisiana* but these species are, in turn, not evenly distributed throughout Europe (Kurtenbach *et al.*, 2001). In Chapter 3 it was observed that *B. afzelii* was found at all three sites in Latvia, however at two of the three sites there was a distinct trend towards reduction in the prevalence of *B. afzelii* over the years investigated. Furthermore in England this was the first reported observation of *B. afzelii* strains in questing ticks but the species was only present at half the sites investigated. In the Scottish highlands, on the other hand, *B. afzelii* was the only species observed in my data and this is similar to previous reports (Ling *et al.*, 2000).

Here, through the use of the MLSA scheme, I investigate the population structure of the three main LB species found in Europe. In Chapter 3 I observed *B. afzelii* in England for the first time but the populations were localised to only a proportion of sites. In Latvia I observed a decrease in *B. afzelii* prevalence at some sites. Here I aim to better understand the findings reported in Chapter 3 through the use of the MLSA scheme. I investigate infections in questing ticks from sites in England, Scotland and Wales and compare the infections with those from questing ticks from Latvia and Germany as well as cultured strains from France. I then compare these European strains to sequence data from the MLST website of strains from China.

I also attempt to identify any changes in the ST frequency in Latvian species between 2002, 2006 and 2007. I show that *B. afzelii* populations in England are different from *B. afzelii* found in Scotland but closely related to those isolated in France. Using these data, I also show that the geographic population structures of European LB species are shaped by their host associations. Using goeBURST and other analyses, I provide evidence that strains of the bird related species, *B. garinii* and *B. valaisiana*, are spatially mixed while, in the rodent related species, *B. afzelii*, spatial structuring of strains is observed. These data strongly support the hypothesis that reservoir hosts are important for spread of LB spirochetes.

Finally I show that the phylogeographic pattern of *B. afzelii* strains is similar to that of other small mammal species and suggest that *B. afzelii* populations may have followed the same post-glacial expansion routes as these small mammal hosts. I also reveal, through the use of pairwise mismatch frequency distributions, that all three species show evidence for population expansion.

5.2 Results

5.2.1 Multilocus Sequence Analysis of European populations

The sequences from the eight MLSA housekeeping genes were analysed from 212 *Borrelia*-positive ticks collected over several years from Britain (n=118), Latvia (n=73) and Germany (n=21) as stated in section 2.9. A full list of strains investigated can be found in Appendix 1, Table A1.1. These data were also compared with sequence data from cultured French strains (n=46) from the Institute Pasteur, France (Appendix 1, Table A1.1). The 72 *B. afzelii* strains were resolved by MLSA into 40 sequence types (STs) (0.56 STs per isolate), the 112 *B. garinii* into 46 STs (0.41 STs per isolate) and the 75 *B. valaisiana* isolates into 26 STs (0.34 STs per isolate). No alleles were found in more than one species and the mean pairwise distances (p-distance) also differed between the species (Table 5.1). *B. garinii* had the highest p-distance of the three species, while *B. afzelii* and *B. valaisiana* showed a similar level of diversity (Table 5.1).

Table 5.1. Mean nucleotide p-distance (π) and mean dN/dS ratio of individual genes and concatenated sequence by country.

	B. afzelii		***B. garinii***		**B. valaisiana**	
	π	**dN/dS**	**π**	**dN/dS**	**π**	**dN/dS**
All samples	0.0019	0.14	0.0078	0.13	0.0018	0.15
English	0.0014	0.14	0.0076	0.11	0.0016	0.14
French	0.0018	0.13	0.0076	0.13	0.001	0.16
German	0.0022	0.14	0.0075	0.12	n/a	n/a
Latvian	0.0013	0.14	0.0081	0.13	0.0025	0.17

5.2.1.1 Distribution of STs between European countries

The majority of STs in *B. garinii* (26/46; 57%) were only found once, but of those STs found at least twice, 13/20 (65 %) were noted in more than one country (Table 5.2, Figure 5.2A). In *B. valaisiana* a similar pattern was observed; 13/26 (50%) of the STs were found only once, but of those found at least twice, 9/13 (69%) were recorded in at least two countries (Table 5.2, Figure 5.2B). For neither *B. garinii* nor *B. valaisiana* was there any suggestion from the data that STs are more likely to be shared between countries in close proximity. For *B. afzelli*, 30/40 (75%) of the STs were found only once, but of those found more than once, 2/10 (20%) were recorded in at least two countries (Table 5.2, Figure 5.2C).

The two *B. afzellii* STs found in more than one country were ST80 (found in Alsace, France and Siebengebirge, Germany), and ST204 (found in Sauerland, Germany and Kemeri, Latvia). The first two sites are in close geographical proximity (Figure 5.1).

The five *B. afzelii* STs in England show a particularly high level of localised clustering. ST164 corresponded to 13 infected ticks, all of which were sampled at Widcombe Hill, Bath. STs 240 and 265 were unique to one and three infected ticks, respectively, from the Exmoor site, and the only example of ST241 was encountered at the New Forest site, Hampshire. Only one *B. afzelii* ST was found at more than one site in England, and the two sites were very close; ST250 was noted in four ticks at Rainbow wood, Bath and in one tick at Warleigh forest, near Bath (Figure 5.1). A higher degree of *B. afzelii* diversity was noted within individual sites in continental Europe than in each English site. For example, all twelve German *B. afzelii* strains investigated were different STs and came from only three different sites. Furthermore, continental *B. afzelii* STs were not always unique to particular sites, as was most often the case in England. For example, the same *B. afzelii* STs were found at different sites within Latvia (albeit mostly exclusively in that country) and a similar pattern was observed with *B. afzelii* STs exclusive to France.

Figure 5.1. Google map of Europe showing tick collection sites. Labels a to v represent the following collection sites; (a) Campus, (b) Bathampton, (c) Widcombe Hill, (d) Rainbow Wood, (e) Warleigh Wood, (f) Inverness, (g) New Forest, (h) Hazeley Heath, (i) Richmond Park, (j) Exmoor, (k) Rhossili Downs, (l) Kemeri, (m) Babite, (n) Jaunciems, (o) Normandie, (p) Meuse, (q) Alsace, (r) Limousin, (s) Auvergne, (t) Kottenforst and Siebengebirge, (u) Lennestadt-Meggen, (v) Wichmar.

Table 5.2. Frequency of all STs where multiple examples were identified and the number of countries the ST was identified in.

Species	ST	Frequency	No of Countries	Countries
B. afzelii	77	4	1	FRA
	80	2	2	FRA, GER
	164	13	1	GB
	165	4	1	LAT
	166	2	1	LAT
	170	3	1	LAT
	204	2	2	GER, LAT
	215	2	1	LAT
	7707	3	1	GB
	250	5	1	GB
B. garinii	82	3	1	GB
	86	13	3	FRA, GB, LAT
	87	5	2	FRA, GB
	88	5	2	FRA, GB
	89	3	2	FRA, LAT
	90	2	2	FRA, LAT
	93	3	2	FRA, GB
	94	2	2	FRA, GER
	163	6	2	GB, LAT
	172	2	1	GB
	173	2	1	GB
	179	2	2	GER, GB
	180	4	2	GER, LAT
	187	9	3	GER, GB, LAT
	188	2	1	LAT
	190	6	2	GB, LAT
	207	6	2	GB, LAT
	243	3	1	GB
	244	3	1	GB
	246	5	1	GB
B. valaisiana	96	10	3	FRA, GB, LAT
	97	13	3	FRA, GB, LAT
	98	2	1	FRA
	99	3	2	FRA, GB
	102	6	3	FRA, GB, LAT
	103	3	2	FRA, GB
	196	2	1	GB
	199	5	2	GB, LAT
	201	5	2	GB, LAT
	203	2	1	LAT
	205	2	1	GB
	211	5	2	GB, LAT
	212	4	2	GB, LAT

Abbreviations for countries are as follows France (FRA), Germany (GER), Great Britain (GB), Latvia (LAT).

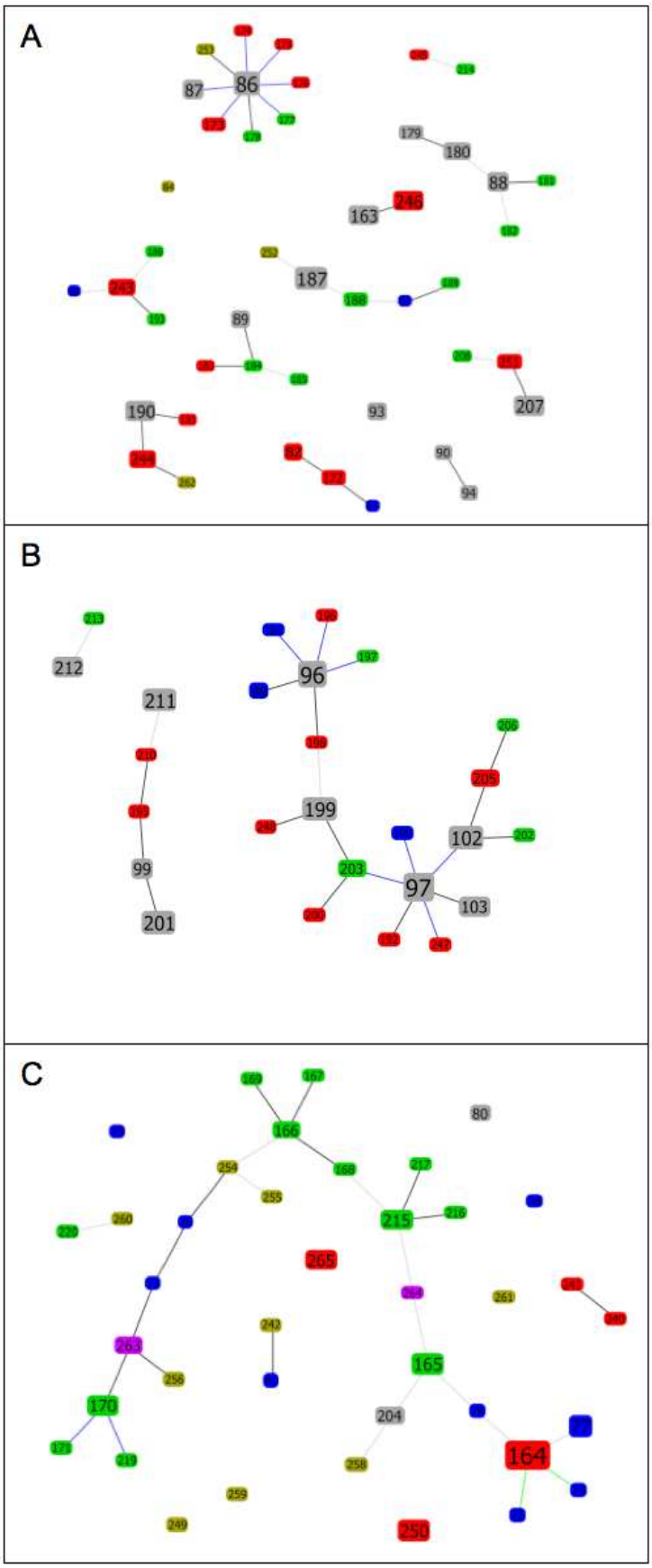


Figure 5.2. goeBURST diagrams based on the multilocus allelic profiles for *B. afzelii*, *B. garinii* and *B. valaisiana*. Each coloured box represents an ST. The colour and size of the boxes corresponds to country and the number of that ST found. STs unique to a particular country were coloured as follows: England, France, Germany, Latvia, Scotland. Those STs that were found in more than one country are . STs connected by black or blue lines are single locus variants (SLV) and STs connected by grey or green lines are double locus variants (DLV). (A) *B. garinii* goeBURST, (B) *B. valaisiana* goeBURST and (C) *B. afzelii* goeBURST.

5.2.1.2 Population differentiation in Europe

I computed pairwise F_{ST} values for the three species, comparing populations from England, Germany, France and Latvia using the concatenated housekeeping gene sequences in ARLEQUIN 3.1. F_{ST} scores estimate the level of differentiation between populations. The output values are defined as being 'low' when between 0.00 and 0.05, 'medium' between 0.05 and 0.25 and 'high' above 0.25 (Freeland, 2005). German samples were not included in the analyses of *B. valaisiana* and *B. garinii*, since no *B. valaisiana* and only nine *B. garinii* strains were found in Germany. Overall, we found low to moderate differentiation in *B. valaisiana* (F_{ST} 0.0477, P= 0.0323) but no statistically significant differentiation in *B. garinii* populations. For a more detailed picture, we then computed pairwise F_{ST} scores for pairs of populations as indicated in Table 5.3. Again, for *B. garinii* no significant differentiation was found. In *B. valaisiana* two of the three pairs of populations showed statistically significant moderate differentiation; these were France/Latvia (F_{ST} 0.0895, P= 0.009) and France/England (F_{ST} 0.0653, P= 0.027).

Table 5.3. Pairwise F_{ST} values of the concatentated housekeeping genes for pairs of populations with values where P is less than 0.05 shown in bold. Analyses that were not completed due to too few strains are indicated as not analysed (N/A).

		England		**France**		**Latvia**	
		F_{ST}	**P value**	**FST**	**P value**	**FST**	**P value**
England	*B. afzelii*						
	B. garinii						
	B. valaisiana						
France	*B. afzelii*	**0.109**	**0.015**				
	B. garinii	0.035	0.058				
	B. valaisiana	**0.065**	**0.027**				
Latvia	*B. afzelii*	**0.364**	**0.000**	**0.22**	**0.000**		
	B. garinii	0.021	0.091	-0.007	0.546		
	B. valaisiana	0.012	0.216	**0.0895**	**0.009**		
Germany	*B. afzelii*	**0.256**	**0.000**	**0.118**	**0.001**	**0.08**	**0.017**
	B. garinii	N/A	N/A	N/A	N/A	N/A	N/A
	B. valaisiana	N/A	N/A	N/A	N/A	N/A	N/A

B. afzelii was found to have the highest overall F_{ST} value of the three species of 0.222 (P= 0.000). All comparisons of pairs of populations were found to be statistically significant but these F_{ST} values showed no significant trend to increase with the geographical distance between the countries (Table 5.3). All the continental European populations, when compared pairwise, showed moderate levels of differentiation ranging from 0.08 (Germany/Latvia) to 0.22 (France/Latvia). The only population pair comparisons that included England fell into the high differentiation category; these were England/Germany (F_{ST} 0.256) and England/Latvia (F_{ST} 0.364) (Table 5.3).

5.2.1.3 Phylogenetic relationship of European **B. afzelii** *populations*

The PhyML tree in Figure 5.3 was created using the concatenated housekeeping genes for all *B. afzelii* STs including any Chinese strains that were found on the MLST database (www.mlst.net). This tree suggests that there is no single common ancestor for all *B. afzelii* strains found in England. However, all English STs and most French STs form a single branch while the majority of Latvian and Scottish STs form the second major branch within the European *B. afzelii* strains (Figure 5.3). The European clade consists of a polytomy with these two major branches as well as several single strains forming minor branches. The goeBURST image in Figure 5.2C shows that the majority of *B. afzelii* STs collected in England are not associated with the main clonal complex of *B. afzelii*, appearing instead as lone STs, meaning that they differ by more than two loci from all Latvian, French, Scottish and German strains included in this study. The PhyML tree suggests that their closest relatives in this dataset are always strains that originated in France or Germany but never in Latvia. This is in contrast to the two STs identified from Scotland, which appear more closely associated with Latvian strains. The only ST from England associated with the large *B. afzelii* clonal complex is ST164 and appears as a founder to the French STs 72, 74, 77.

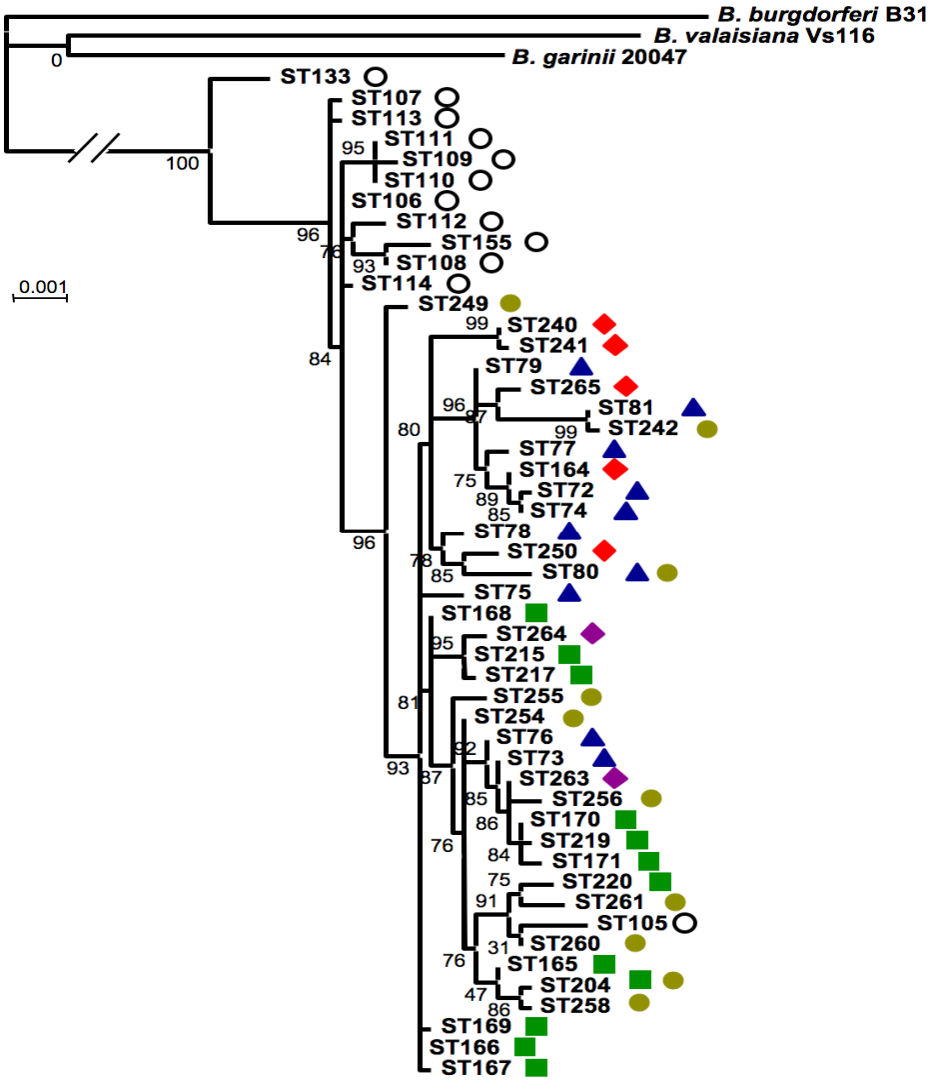


Figure 5.3. PhyML phylogenetic inference of the concatenated housekeeping genes of *B. afzelii* STs. The symbols indicate the country the ST was found in; ▲ France, ◆ England, ■ Latvia, ● Germany, ◆ Scotland and ○ China. The tree is rooted with three LB species; *B. garinii, B. afzelii* and *B. burgdorferi* but the branch length is not to scale as indicated by the slashes.

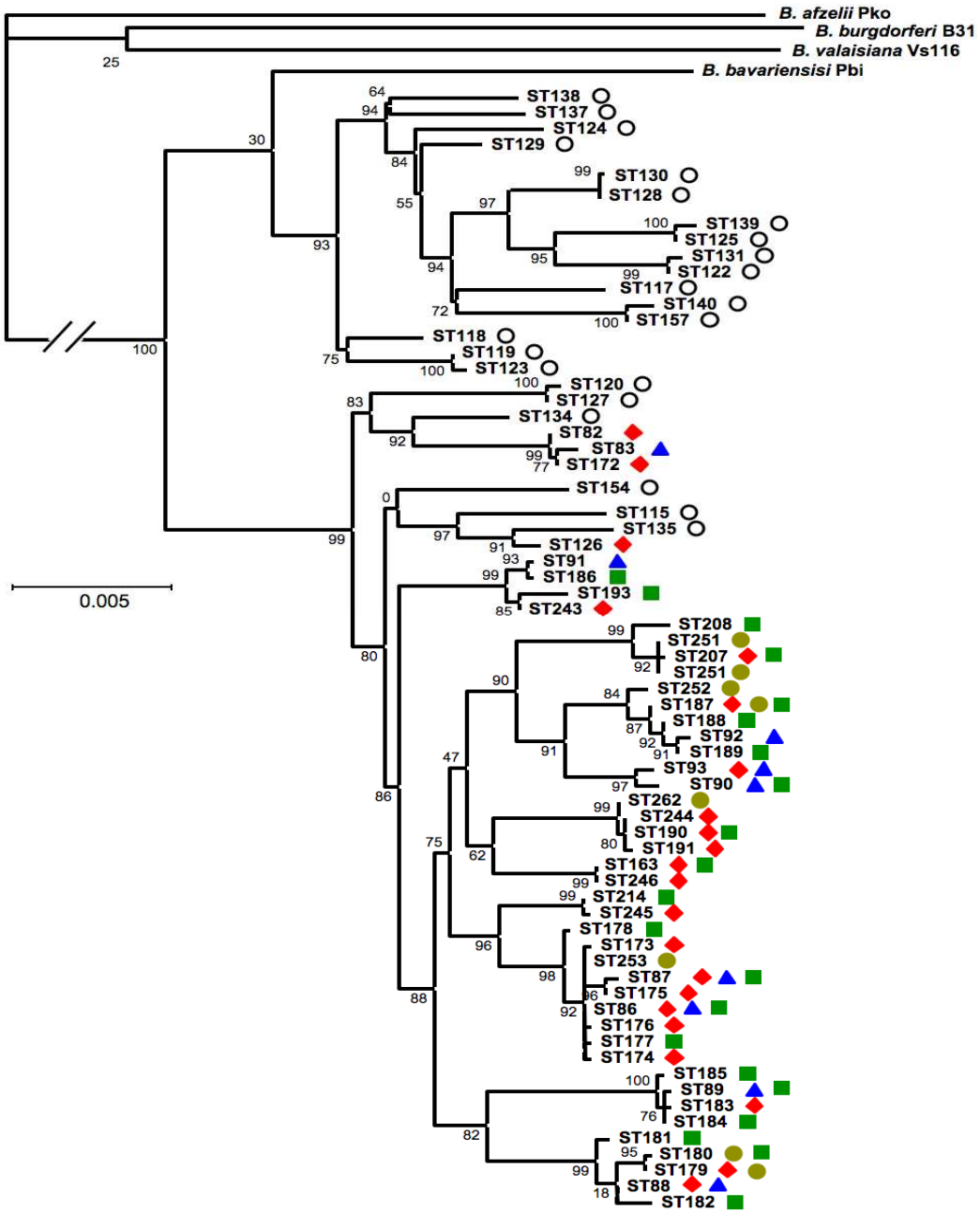


Figure 5.4. PhyML phylogenetic inference of the concatenated housekeeping genes of *B. garinii* STs. The symbols indicate the country the ST was found in; ▲ France, ◆ England, ■ Latvia, ● Germany and ○ China. The tree is rooted with three LB species; *B. garinii, B. afzelii* and *B. burgdorferi* but the branch length is not to scale as indicated by the slashes.

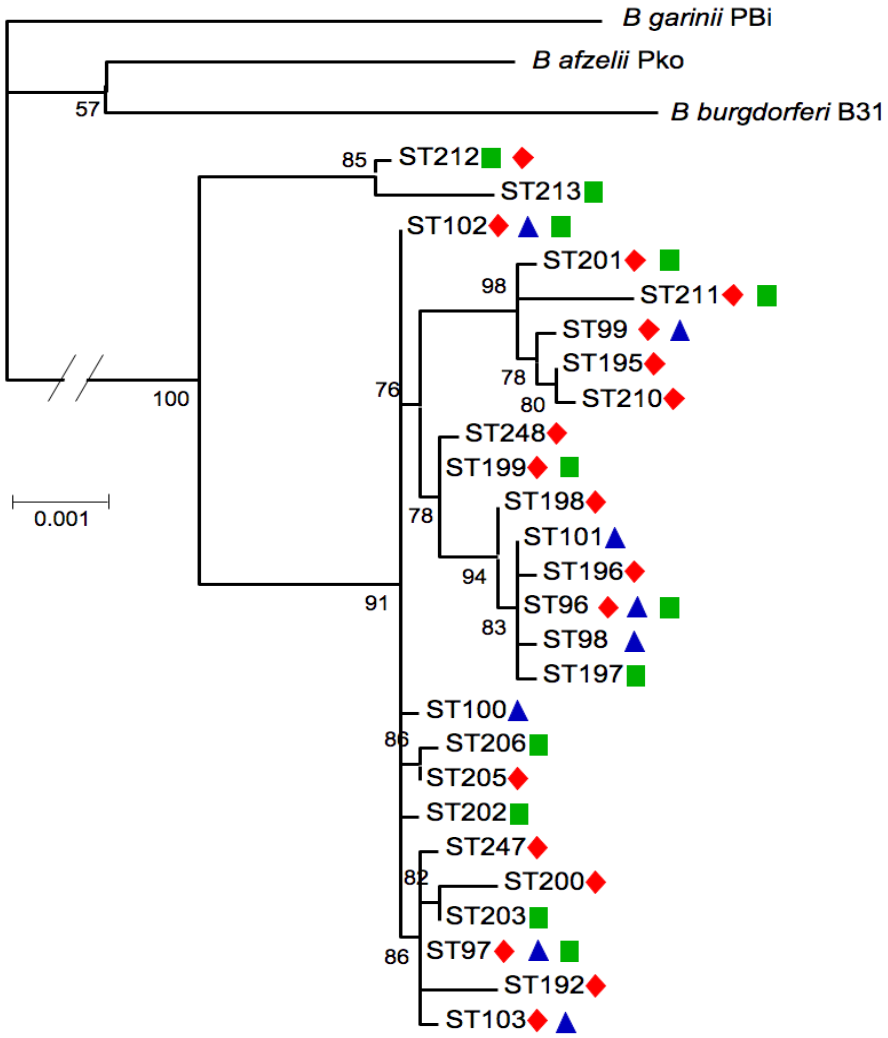


Figure 5.5. PhyML phylogenetic inference of the concatenated housekeeping genes of *B. valaisiana* STs. The symbols indicate the country the ST was found in; ▲ France, ◆ England, ■ Latvia. The tree is rooted with three LB species; *B. garinii, B. afzelii* and *B. burgdorferi* but the branch length is not to scale as indicated by the slashes.

Similar to the goeBURST diagrams, the PhyML trees for *B. garinii* and *B. valaisiana* show strains obtained from the different European countries do not cluster in accordance with geographic location, but strains from different countries appear in the same cluster/terminal nodes. The mixing of populations apparent from these phylogenetic relationships would be expected to have arisen from a high level of migration (Figures 5.4 and 5.5).

5.2.2 Pairwise Mismatch Distributions of European Species

It has been shown that frequency distribution of pairs of sequence mismatches can leave signatures of past population demographic events (Rogers and Harpending, 1992). Here, I examined the pairwise mismatch distributions of the concatenated gene sequences in the three LB group species (*B. afzelii*, *B. garinii* and *B. valaisiana*) for evidence of past spatial

and population expansions by comparing them with predicted distributions obtained under models of constant population size, population expansion, and spatial expansion. Pairwise comparisons of all European strains from *B. afzelii*, *B. garinii* and *B. valaisiana* were made using Arlequin (Excoffier *et al.*, 2005) as described in section 2.11.3 and the frequency of pairwise mismatches were plotted on graphs in figures 5.6 to 5.8. Model frequency distributions were calculated for spatial and demographic population expansions, using Arlequin and constant population, using DNASP (Librado and Rozas, 2009) (Section 2.11.3). These model frequency distributions were also plotted in figures 5.6 to 5.8.

The observed pairwise mismatch data for all three species supported the spatial expansion model but there was a high degree of variation in tau (τ) values between species (Table 5.4). In *B. garinii* and *B. valaisiana* populations demographic expansion could not be excluded (P = 0.1 and P=0.55, respectively) while, *B. afzelii* was the only species where it was possible to exclude demographic population expansion (P = 0.02). All three species appeared visually less similar to the constant population model as apposed to the two expansion models (Figures 5.6 to 5.8).

Table 5.4 P-values calculated by Arlequin showing probability that the observed data does not fit the demographic or spatial model simulated data. Models that were not excluded are shown in bold. Arlequin estimates of τ are also shown in the cases where the model was not excluded.

	Demographic expansion		**Spatial expansion**	
	τ	P value	τ	P value
B. afzelii	n/a	0.02	10.2	**0.23**
B. garinii	47.5	**0.1**	39.5	**0.62**
B. valaisiana	5.9	**0.55**	4.3	**0.78**

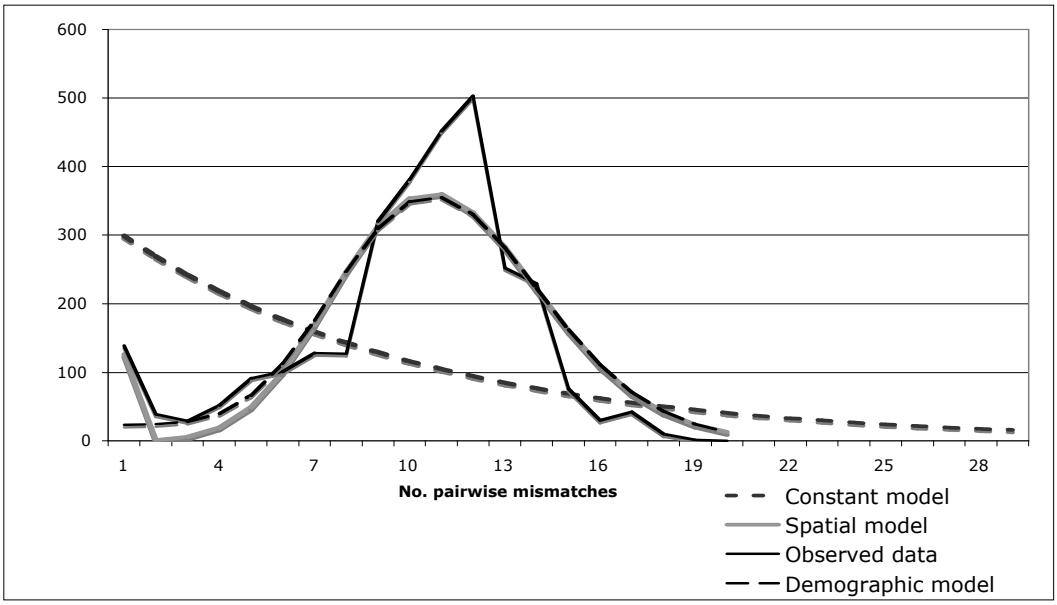


Figure 5.6. *B. afzelii* mismatch distribution. Frequency graph showing observed paiwise nucleotide differences for the concatenated housekeeping genes (black line) and the model nucleotide differences for three population demographic scenarios; a constant population, a spatial population expansion and demographic population expansion.

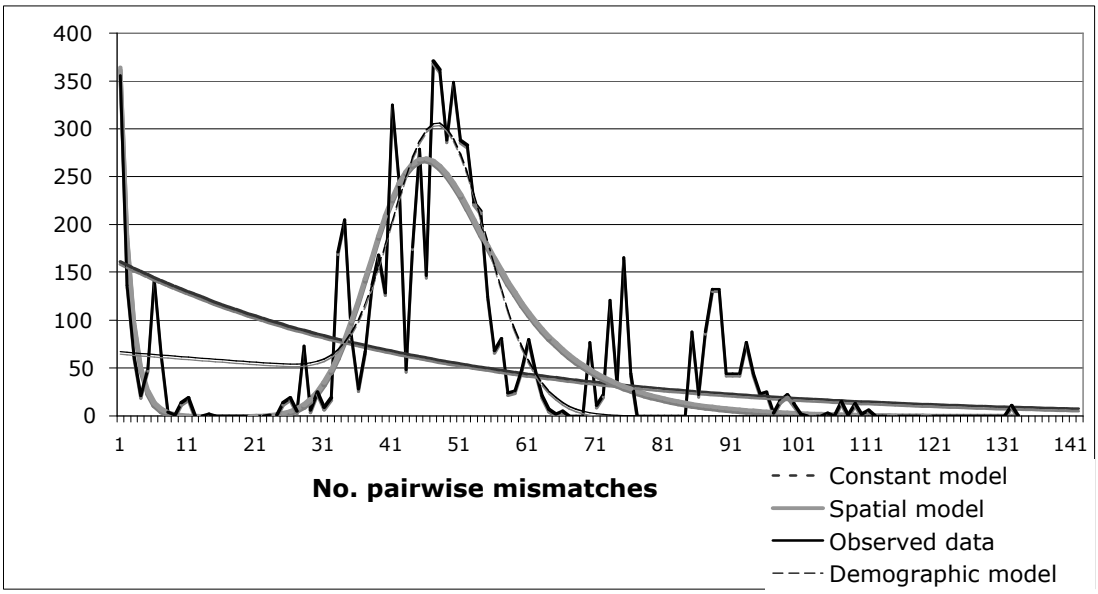


Figure 5.7. *B. garinii* mismatch distribution. Frequency graph showing observed pairwise nucleotide differences for the concatenated housekeeping genes (black line) and the model nucleotide differences for three population demographic scenarios; a constant population, a spatial population expansion and demographic population expansion.

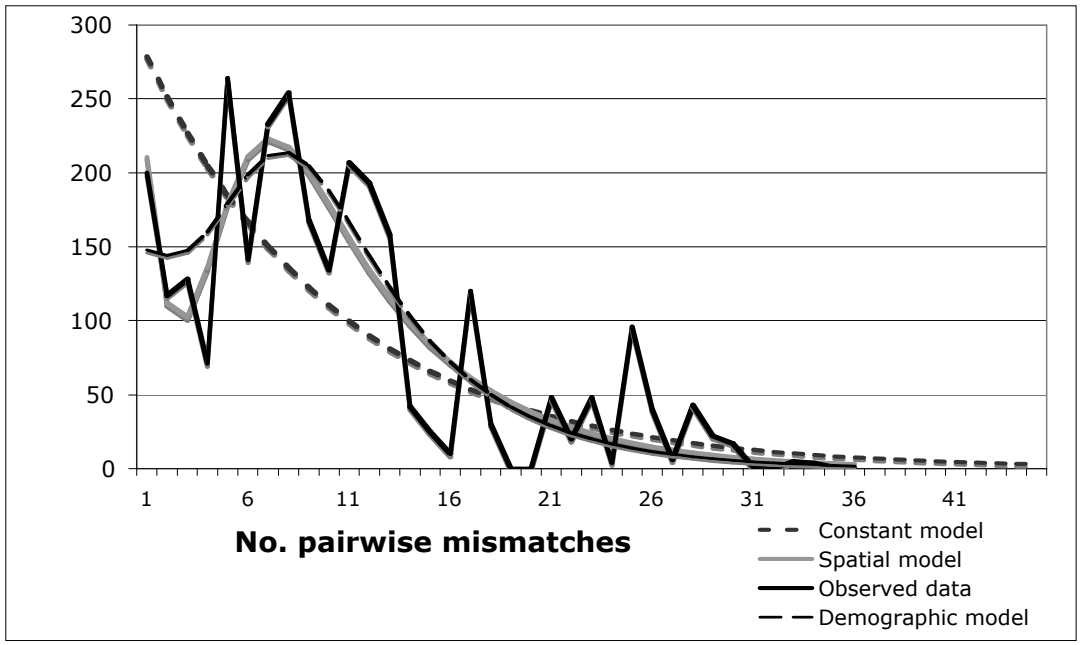


Figure 5.8. *B. valaisiana* mismatch distribution. Frequency graph showing observed pairwise nucleotide differences for the concatenated housekeeping genes (black line) and the model nucleotide differences for three population demographic scenarios; a constant population, a spatial population expansion and demographic population expansion.

5.2.3 Comparison of European and Chinese populations

The European data for *B. afzelii* and *B. garinii* was then compared with Chinese strains available on the MLST.net database. Chinese strains assigned to *B. garinii* and *B. afzelii* were extracted from the database and included in the phylogenies in Figures 5.3 and 5.4. There was no overlap between STs found on the different continents for either *B. garinii* or *B. afzelii*. The majority of Chinese *B. garinii* strains appeared to be related to *B. garinii* serotype 4, which, is rodent associated and has recently been redefined as *B. bavariensis* (Margos *et al.*, 2009). There were only seven strains that appear to be true *B. garinii* strains forming six different STs (Figure 5.4). These true *B. garinii* strains did not form a single monophyletic group but instead fall on two branches that also included European strains albeit on relatively long branches while the majority of European STs did form a large monophyletic group. However, the limited number of true *B. garinii* strains makes it difficult to draw further conclusions.

Fifteen *B. afzelii* strains from Chinese questing ticks were available on the database and all but one fell on several branches separate from European strains. All the European strains formed a monophyletic group however there was one Chinese strain that is within this group suggesting a possible migration event. The F_{ST} value comparing the Chinese and European samples was 0.404 (P=0.0000), which was substantially higher than the comparison between European countries (F_{ST}=0.222). The pairwise comparisons between

China and the individual European countries ranged from 0.366 to 0.422 (P=0.0000) (Table 5.5) and were only slightly above the range of pairwise comparisons within Europe.

Table 5.5 Pairwise F_{ST} values of the housekeeping genes for pairwise comparisons of European populations to the Chinese popluation.

	China	
	F_{ST}	P-value
England	0.391	0.0000
France	0.366	0.0000
Germany	0.389	0.0000
Latvia	0.422	0.0000
Europe	0.404	0.0000

5.2.4 Temporal study of Latvian sites

The eight housekeeping genes were amplified for 45 infected Latvian ticks from 2002 from the three sites discussed in Chapter 3, section 3.2.2.1. These infections consisted of 20 *B. afzelii*, 17 *B. garinii*, 4 *B. burgdorferi*, 3 *B. valaisiana* and 1 *B. lusitania* (Appendix 4, Table A4.2). It was considered that *B. afzelii* and *B. garinii* had sample sizes large enough to compare the proportion of STs between years. For these two species, the samples collected in 2002 were compared to the Latvian samples collected in 2006 and 2007 that have already been discussed (Section 5.2.1).

In *B. afzelii* no single dominant ST was observed in any year and the same STs did appear in different years. STs that were more common in one year were not necessarily common in other years but the samples sizes were too small to identify trends of particular STs to any level of certainty. *B. garinii* showed a greater trend of having STs unique to single years with approximately half of all isolates being STs unique to the year of collection (Figure 5.10). However, as many of these were single STs (only found once in the Latvian tick collection) it was considered that grouping of isolates by ST maybe too specific to understand trends between years so STs were then grouped into clonal complexes. The clonal complexes were defined as single locus variants of a putative founder strain when considering *B. afzelii* STs. For *B. garinii* STs both single and double locus variants were grouped to form clonal complexes due to the greater diversity of strains in the *B. garinii* species (Table 5.6).

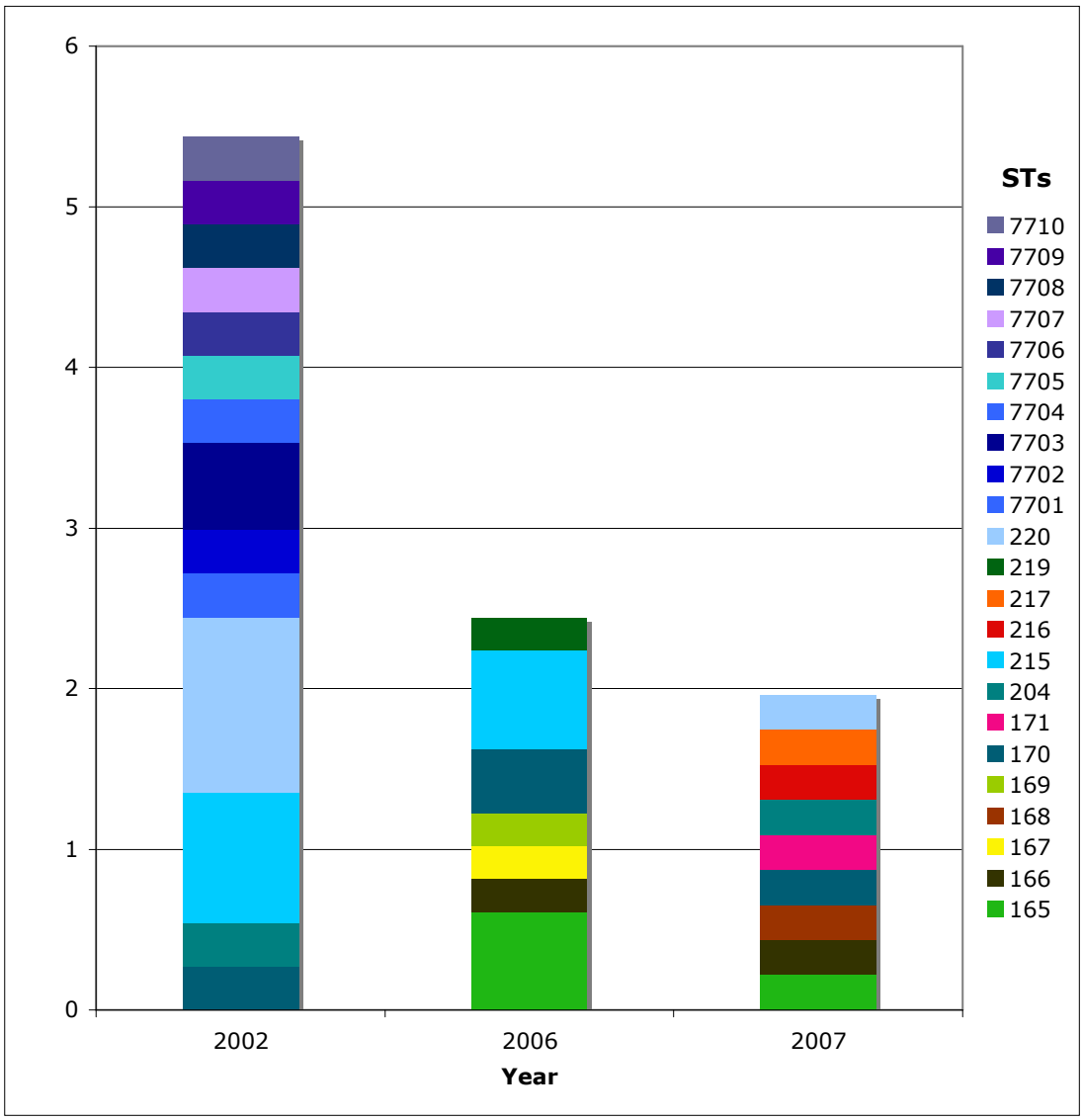


Figure 5.9 Temporal distribution of *B. afzelii* STs. Percentage of ticks collected in 2002, 2006 and 2007 infected with STs of *B. afzelii*. STs were coloured in chronological order meaning those first found in 2002 were coloured in shades of blue, those found first in 2006 were coloured shades of green/yellow and those found only in 2007 were shaded in reds/pinks.

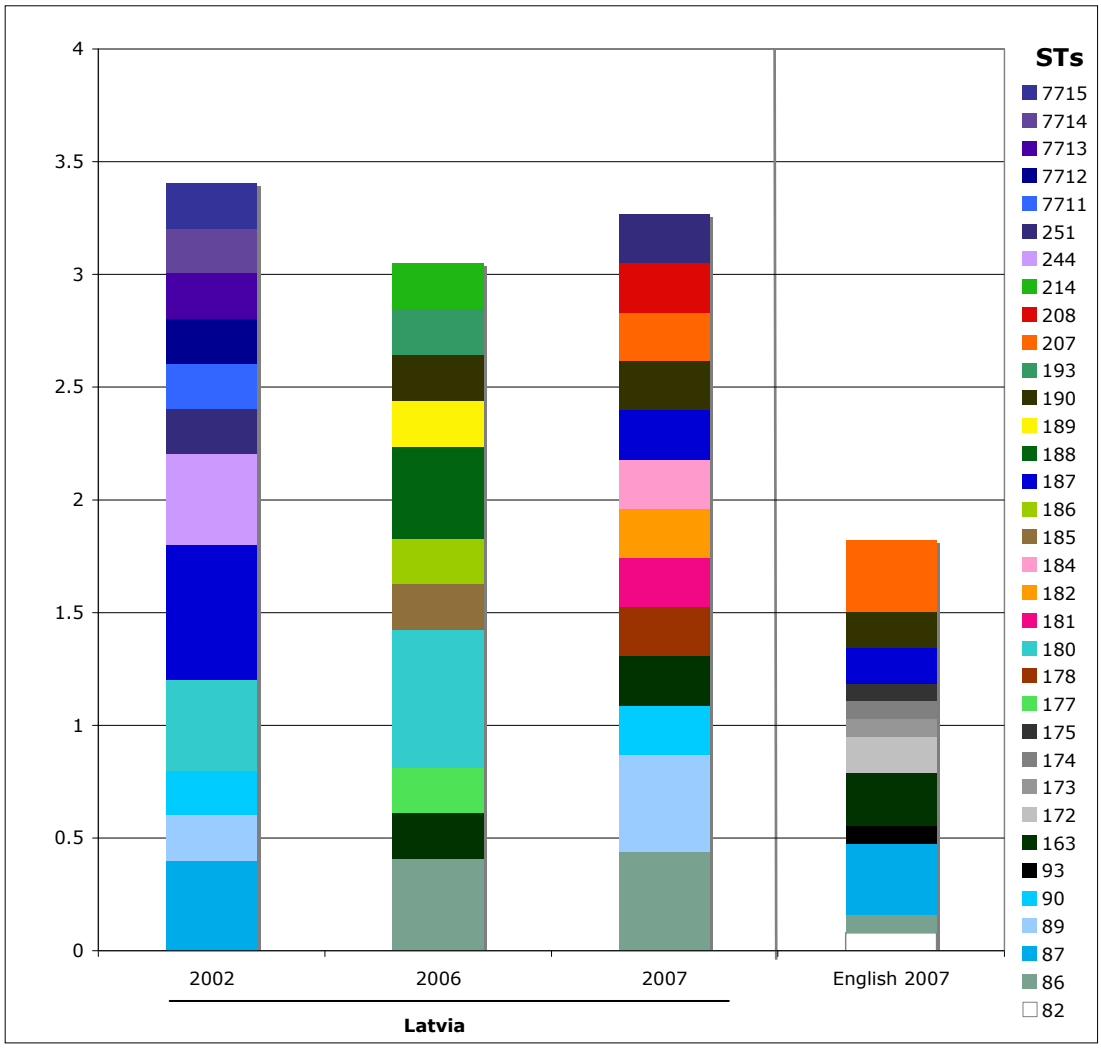


Figure 5.10. Temporal distribution of *B. garinii* STs. Percentage of ticks from 2002, 2006 and 2007 infected with *B. garinii* STs. STs were coloured in chronological order meaning those first found in 2002 were coloured in shades of blue, those found first in 2006 were coloured shades of green/yellow and those found only in 2007 were shaded in reds/pinks. Those unique to English sites were coloured grey.

The clonal complex frequency graphs in Figures 5.11 and 5.12 showed most clonal complexes were found in all years in both *B. garinii* and *B. afzelii* but there was variation in how common each clonal complex was between the years. There was also no dominant clonal complex in either *B. garinii* or *B. afzelii* in any of the years. Interestingly, there appears to have been a greater loss of apparent singleton STs from the *B. afzelii* population in Latvia compared to those that are within common clonal complexes (this is most notable when comparing the percentage of infected ticks in 2002 in Figures 5.9 and 5.11). However, a greater sample size would be needed to confirm this finding as there are too few samples to analyse this trend statistically.

Since *B. garinii* did not show spatial structuring when samples from a wide geographic area were compared (see 5.2.1.1), STs and clonal complexes from English ticks were included for comparison. English 2007 STs did not show greater similarity to Latvia

2007 than to 2002 and 2006 (Figure 5.10). It did, however, contain one clonal complex that was not identified in Latvia in any year. In *B. afzelii* all years have 4 clonal complexes and in *B. garinii* all years have 5 or 6 clonal complexes (including England 2007) and when one clonal complex is absent in a particular year it is replaced by a different clonal complex meaning the overall number of clonal complex groups remains constant. However, more sampling years would be necessary to confirm this finding.

Table 5.6 Shows STs that make up the clonal complexes and which in years the STs were observed.

Species	Clonal Complex	STs	Years observed
B. afzelii	166	166	2006, 2007
		167	2006
		168	2007
		169	2006
	165	165	2006, 2007
		7701	2002
	170	170	2002, 2006, 2007
		171	2007
		219	2006
		7704	2002
	215	215	2002, 2006
		216	2007
		217	2007
	7703	7703	2002
		7710	2002
B. garinii	86	86	2006, 2007
		87	2002
		177	2006
		178	2007
	180	180	2002, 2006
		7715	2002
	182	181	2007
		182	2007
	184	89	2002, 2007
		184	2007
		185	2006
		7711	2002
	188	187	2002, 2007
		188	2006
		189	2006
		7712	2002
	244	190	2006, 2007
		244	2002
	251	207	2007
		208	2007
		251	2002, 2007
		7714	2002

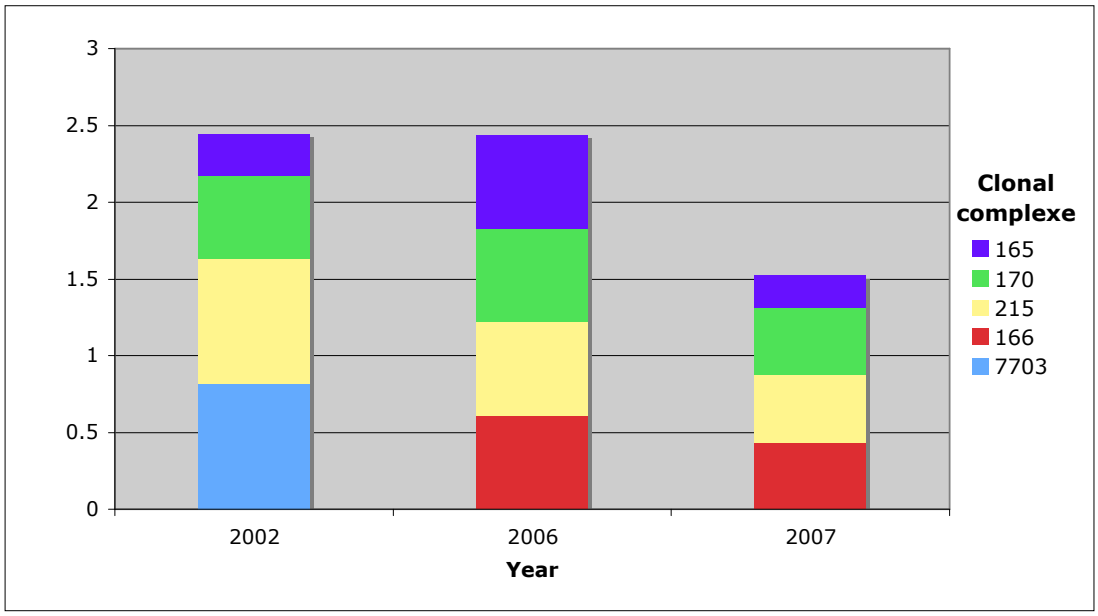


Figure 5.11. Temporal distribution of *B. afzelii* CC groups. Percentage of ticks infected with *B. afzelii* clonal complex groups during 2002, 2006 and 2007 in Latvia.

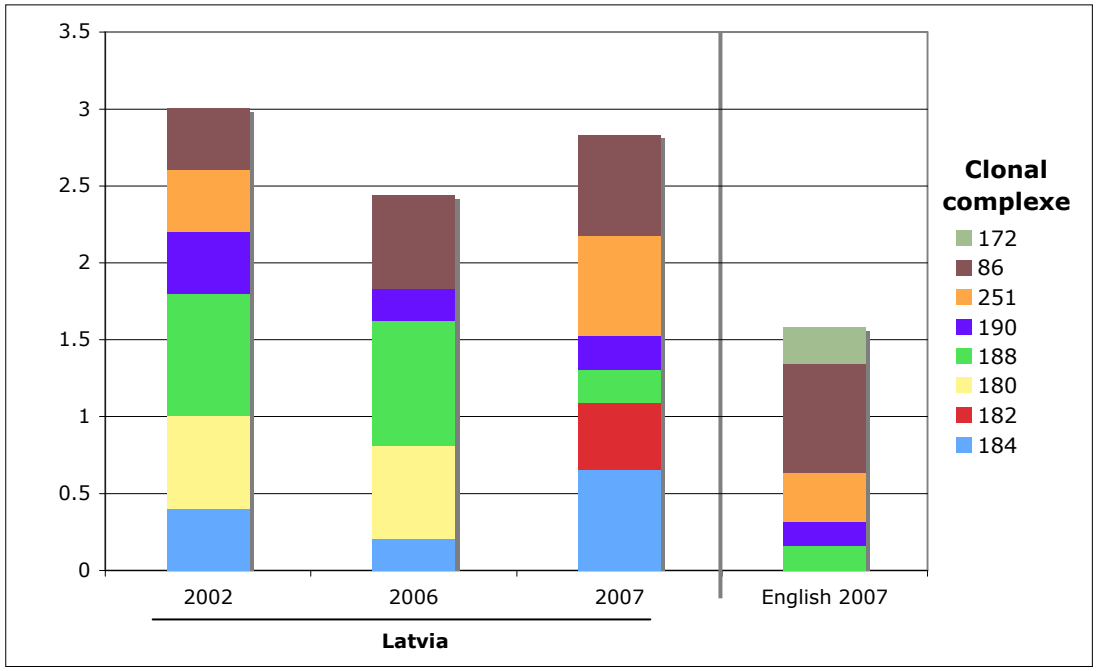


Figure 5.12 Temporal distribution of *B. garinii* CCs. Percentage of ticks from Latvia collected in 2002, 2006 and 2007 infected with *B. garinii* clonal complex groups and from England collected in 2007.

5.2.4.1 Spatial Structure of B. afzelii population in Latvia

The Latvian *B. afzelii* populations appeared to show little ST bias between the three years so data from the three years was pooled to compare the distribution of *B. afzelii* STs from different Latvian sites. Figure 5.13 shows that Jaunciems and Jurmala share many of the same *B. afzelii* STs but that Babite appears to have many more STs unique to this site. Figure 5.14 depicts the clonal complexes found at the different sites and shows a similar

pattern which suggests that there may be limited migration between Babite and the other two sites. Four of the nine strains unique to Babite form a clonal complex unique to Babite (clonal complex 7703). Two of the nine strains have ST 168 which was also found at the Jaunciems site. The suggestion that there was more migration between Jaunciems and Jurmala compared with Babite is surprising as the two sites are separated by the city of Riga (Figure 5.1). While, Jurmala and Babite are the only pair of sites not separated by the city and yet have no shared STs or clonal complexes.

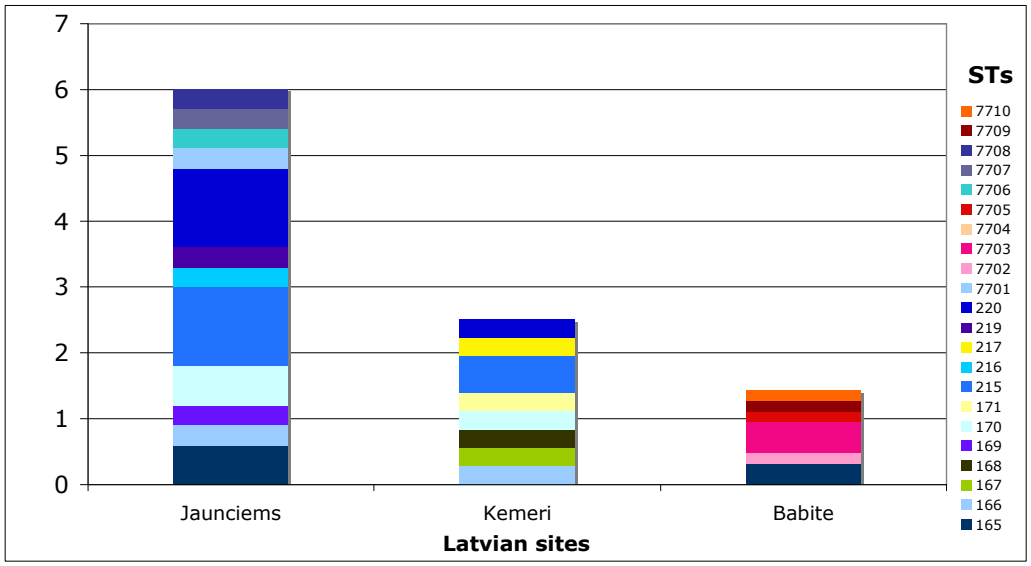


Figure 5.13 Percentage of ticks infected with *B. afzelii* STs collected at Jauniems, Jurmala and Babite. STs were coloured starting with those from ticks collected in Jaunciems (shades of blue), then those from ticks collected in Jurmala (shades of green/yellow) and finally from Babite ticks (shades of red/pink).

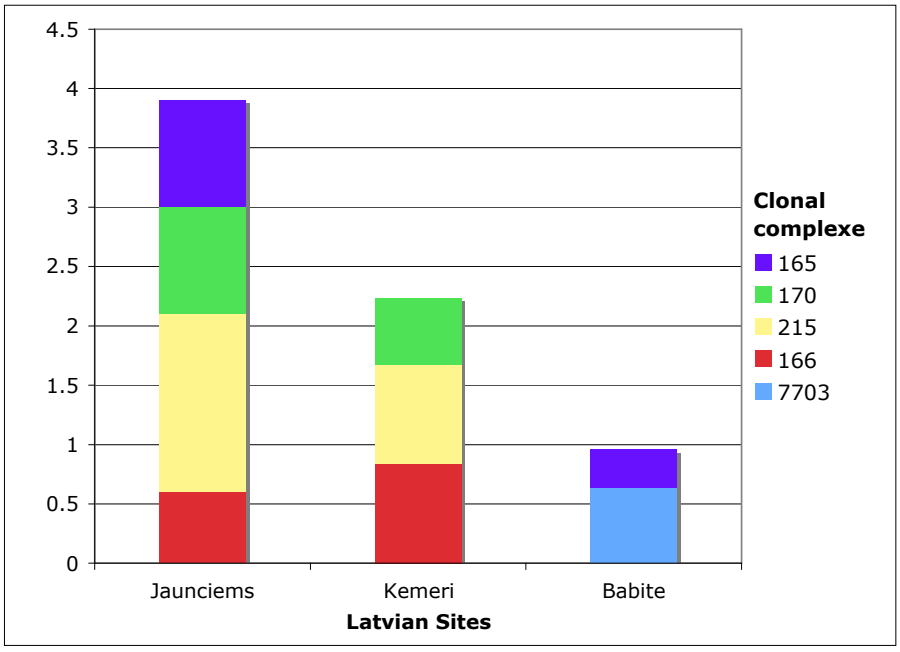


Figure 5.14 Percentage of ticks infected with *B. afzelii* clonal complex groups at the Latvian sites Jaunciems, Jurmala and Babite

5.3 Discussion

5.3.1 Phylogeographic Structuring of LB Species

Host specialisation is a key process in the ecology and evolution of tick-borne zoonotic diseases. In this study I aimed to further our understanding of the impact of host association on the spread of zoonotic vector-borne pathogens by analysing three species of the LB group of spirochetes that are specialized to either avian or rodent hosts. MLSA on housekeeping genes has revealed differences in the level of geographic structuring of populations of LB species that are consistent with patterns of migration of their different vertebrate hosts. Both of the bird-related species investigated, *B. valaisiana* and *B. garinii*, showed evidence of spatial mixing of STs between countries, whilst the rodent-related *B. afzelii* showed evidence of differentiation of populations from each of the four countries. This differentiation was pronounced to the extent that only two *B. afzelii* STs were found in more than one country.

This finding was statistically supported using an F_{ST} test for pairwise differentiation between populations. Interestingly, while in *B. garinii* no significant differentiation of populations was found in different countries, suggesting entirely free movement of strains, *B. valaisiana* showed low to moderate differentiation, suggesting there is not complete homogenization of *B. valaisiana* strains within Europe. This was surprising because both species appear to be transmitted by similar species of avian hosts (Dubska *et al.*, 2009, Taragel'ova *et al.*, 2008). However, these results suggest that subtle ecological differences may exist between these species and it has been known for a long time that, apart from transmission cycles in terrestrial birds, *B. garinii* populations are also maintained by seabirds and their associated tick, *I. uriae* (Bunikis *et al.*, 1996, Larsson *et al.*, 2007, Olsen *et al.*, 1995). It is interesting to note that Comstedt and colleagues (Comstedt *et al.*, 2009) reported an overlap of marine and terrestrial *B. garinii* populations, though whether or not this may play a role in the population structure observed in the present study remains speculative, since no *B. garinii* from marine transmission cycles have been analysed by MLSA to date.

Of the three species analysed in the present study, *B. garinii* was found to be the most diverse. The finding that *B. garinii* showed a higher genetic diversity than *B. valaisiana* has been reported previously (Margos *et al.*, 2009). This may also contribute to subtle ecological differences in host associations, and therefore to the differences in population patterns observed here. In the pairwise F_{ST} comparisons, the only population pair showing no significant differentiation in *B. valaisiana* was England/Latvia, which is

the most geographically distant pair. Currently, it is not known whether the differentiation observed between the LB strains from France and England is due to bias created by the culturing of the LB strains from France. Further sampling of environmental LB strains from France would be required to confirm these findings.

The *B. afzelii* F_{ST} score was markedly higher than *B. garinii* or *B. valaisiana*, suggesting that movement of strains between geographic regions is more restricted. The higher F_{ST} values observed between England and some European countries suggest that the English Channel is acting as a barrier to the movement of *B. afzelii* strains between Great Britain and continental Europe. The Chinese *B. afzelii* strains were shown to be significantly different from the European strains to a high degree. However, the F_{ST} scores did not increase proportional to distance, producing only slightly higher scores than the comparisons between European countries. This may be an artefact of the limited diversity within the *B. afzelii* species. However, it may suggest that distance may not be the main cause of geographic separation. One *B. afzelii* strain falls within the European group suggesting that there may be rare cases of migration between East and West but it is hard to speculate the method behind such a migration event.

No European *B. garinii* STs were identified in the six Chinese *B. garinii* STs and these Chinese STs showed substantial divergence from European strains indicated by the long branches joining them to the closest European relative (Figure 5.4). This suggests that the free movement of strains observed within Europe, probably instigated by passarine birds (Taragel'ova *et al.*, 2008), may not be maintained over such a large distance. While, migratory sea birds have been known to move *B. garinii* strains over large distances, such as between Europe and the east coast of the US, the isolated nature of the marine bird colonies means the *B. garinii* strains have not spread throughout the US (Smith *et al.*, 2006). Thus, the role that migratory birds play in the movement of the *B. garinii* may be limited by the movement of strains between woodland populations (vectored by *I. ricinus*, *I. scapularus* and *I. persulcatus*) and sea bird related populations (vectored by *I. uriae*) of *B. garinii* (Comstedt *et al.*, 2009). However, more Asian *B. garinii* samples are required to confirm this and access the level of migration of *B. garinii* between East Asia and Europe.

In Latvia it is more common for the same *B. afzelii* ST to be found in different sites compared to England where different sites tend to be associated with specific sequence types. The three Scottish strains included in this study appear to be more closely related to Latvian STs than to English STs (Figure 5.3), suggesting that there is limited, or potentially no, migration between north and south in the UK. This is interesting in light of a study by Searle and colleagues (Searle *et al.*, 2009) investigating three species of small

mammals (including the field vole *Microtus agrestis*, bank vole *Myodes glareolus*, and pygmy shrew *Sorex minutus*) in Great Britain. In each case, two geographically distinct phylogroups were found. All the species showed a clear north/south divide between the phylogroups, which occurred in northern England, in the case of *M. agrestis*, and as far north as central Scotland in the case of *S. minutus*. The marked differentiation between English and Scottish *B. afzelii* samples may therefore be a result of limited north-south rodent migration.

Taken together, the findings reported here strongly support the hypothesis that the movement and spread of LB species is dependant on the host, not the vector, but further investigations on a finer genetic scale, and much denser sampling, are required to fully understand the subtle ecological differences between hosts and to reconstruct fully the migration patterns of the different LB species. Given that the migration of some LB species is limited by the propensity for their vertebrate hosts' ranges to shift, landscape genetic analysis would be an appropriate approach to determine barriers to migration (Manel *et al.*, 2003). This is particularly important for understanding the migration and emergence patterns of *B. afzelii*, which is associated with rodent hosts. Such future investigations would be facilitated by i) identification of the rodent hosts of *B. afzelii* in England and ii) the use of neutral genetic markers such as SNP analyses to identify the fine scale population structure.

5.3.4 Proposed post-glacial expansion of *Borrelia* species

Several studies of potential host species of *B. afzelii* including shrew and vole species have also observed phylogeographic structuring of populations across Europe. Similar to my findings they have found little evidence of a trend of increasing diversity with geographic distance, and have instead attributed the phylogeographic patterns to the ancestral refugia of the populations during the last glacial maximum (LGM) (Fumagalli *et al.*, 1996, Heckel *et al.*, 2005, Hewitt, 1999, Taberlet and Bouvet, 1994, Taberlet *et al.*, 1994). Hewitt (1999) summarises some of these species suggesting two glacial refugia, an Iberian and an East Baltic one for both *Sorex araneus* (common shrew) and *Crocidura suaveolens* (white-toothed shrew) as well as for another small mammal, *Erinaceus spp* (hedgehog). He suggests that populations from the two refugia have since spread northward meeting potentially around Germany or the Czech Republic (Figure 5.15). My data of European *B. afzelii* strains roughly separate them into two clades (Figure 5.3) which could represent populations maintained by the two refugia populations mentioned above. In the phylogeny one branch contains the majority of French and English strains while the other contains

most of the Latvian and Scottish strains suggesting the branches may have originated in the Iberian and Baltic refugia, respectively. The mixture of German strains between the two may suggest that the two refugia populations have met and are mixing in Germany. The p-distance values for *B. afzelii* populations also support these proposed post glacial expansion routes (Figure 5.15) as both of the most northerly countries included in my study have the lowest p-distance values. These were Latvia (0.0013) and England (0.0014) which suggests that they may have been colonised more recently (Page and Holmes, 1998).

Figure 5.15 Proposed post-glacial migration routes for three small mammal species. The image was modified from Hewitt (1999) based on fossil and molecular data.

The two main branches of European *B. afzelii* strains form part of a polytomy with several strains falling outside the main branches. This is consistent with the view that *B. afzelii* have had additional refugia during the LGM, such as potentially a third Italian refugia (Figure 5.15), which reflect the adaptation of this species to many different hosts. This may have caused the more fragmented picture in the *B. afzelii* phylogeny compared to individual small mammal species phylogenies such as the one reported by Heckel and colleagues (2005).

Further to this, the pairwise mismatch distribution for each of the three species (Figures 5.6 to 5.9) suggests that all three species examined have undergone a population expansion. This means that although no geographic structure was observed in either bird related species, they too may have retreated with vector and hosts during the LGM. Interestingly, Hoen and colleagues (Hoen *et al.*, 2009) investigating *B. burgdorferi* population structure in the US suggested an ancient population expansion of *B. burgdorferi,* (in the range of 1000s of years ago) from refugia located on Long Island, New York and in the Mid West.

The major caveat to this theory is that the tau estimates by Arlequin for the three species differed by nearly ten-fold between the highest estimate in *B. garinii* and the lowest in *B. valaisiana* (Table 5.4). Tau estimates the number of mutations that have occurred since the spatial expansion and if the mutation rate and generation time of the

different species is assumed equal, this suggests that the expansions occurred at very different times. The generation time of *Borrelia* species is difficult to estimate due to the lifestyle of the vector-borne zoonoses (Hoen *et al.*, 2009). It is likely to vary greatly throughout the infection cycle where it has been shown that there are very few genome replication cycles of the spirochaetes when in dormant ticks but then it may increase to a generation time of 4hrs during nymphal attachment to the host (De Silva and Fikrig, 1995). The variation in tau values may suggest that the species began to expand at different times. Considering an ice age is a highly gradual process and that *B. garinii* has a broader host range than *B. valaisiana* that includes sea birds found in the Arctic region (Larsson *et al.*, 2007), it is possible that the species expansions began at different points in the ice age. However, the extreme difference in tau values may certainly represent expansions that were caused by separate ancient events but it is difficult to speculate what these may have been. Both possibilities again highlight potential differences in the niches occupied by *B. valaisiana* and *B. garinii*.

5.3.1 The discovery of *B. afzelii* in England, its distribution and its origin

As discussed in Chapter 3, I have reported for the first time *B. afzelii* in questing ticks from England. Just over half the sites were found to be supporting *B. afzelii* strains, and here I show, through the use of the MLSA scheme, that the infections were highly focal, four of the five STs from England each were found exclusively and often repeatedly at a single site. Although it is possible that isolated *B. afzellii* populations have been maintained in England for long periods of time, and may have diverged from each other by genetic drift, I consider it more likely that the genetically distinct localised populations represent independent and fairly recent introductions from outside the UK. This view is supported by the PhyML tree, which reveals that the English STs are polyphyletic, hence there is no single common ancestor unique to English STs (Figure 5.3). However, at present it is only possible to speculate on these strains' geographical origins. The goeBURST reveals that most English STs are not linked to the major clonal complex of the species (Fig. 2C), suggesting they have not recently spread from France, Germany or Latvia. The exception to this is ST164, which was observed at the Widcombe Hill site in England and is assigned as founder to three French STs in Fig. 2C. This infers a French origin of this strain, although the exact ST was not present in our sample from France. In general, the PhyML tree in Figure 5.3 suggests that the English STs are more closely related to French and German STs than to the Latvian and Scottish STs thus supporting the

theory that these strains may have, in the distant past, originated from an Iberian glacial refuge site.

Genomic regions with more genetic variation are required, such as whole genome intergenic SNP analysis, to allow for an estimate on the period since *B. afzelii* introduction to the English sites. Furthermore, a more extensive survey of European sites may reveal the origins of the English *B. afzelii* strains and also shed light on potential methods of entry for these rodent-associated strains.

5.3.2 Fine scale genetic structuring of *B. afzelii* strains

Sites supporting *B. afzelii* strains in England demonstrated a high degree of structuring with only one heterogeneous population found out of the three sites where multiple *B. afzelii* infected ticks were collected and furthermore, only one ST was identified at more than one site. This situation appears unique to English sites. However, Latvian sites also showed a trend of genetic structuring although to a lesser extent. The majority of STs found in Babite were unique to this site but there appeared to be genetic exchange between the other two sites (Figure 5.13). Interestingly, fine scale structuring has been observed in vole species where it has been illustrated that natural and man made barriers such as large rivers and highways reduce the rates of migration between populations of *Clethrionomys glareolus* (bank voles) (Gerlach and Musolf, 2000). Curiously, there appears to be more genetic exchange between the two Latvian sites, Jaunciems and Jurmala, although these are separated by the city of Riga and the river Daugava (Figures 5.13 and 5.14). It is currently unknown why this should be the case and it seems that more samples are required to confirm and measure the genetic exchange between sites. Fine scale structuring in the absence of barriers has also been observed in small mammal populations such as common voles (*Microtus arvalis*) were it has been suggested that social structure within populations means that dispersal from local groups is rare (Gerlach and Musolf, 2000, Schweizer *et al.*, 2007). This mirrors what is seen in *B. afzelii* strains at the Babite site in Latvia and the English sites Widcombe hill and Rainbow wood (Figure 5.1) where minimal migration has been observed even in the absence of large barriers and thus further suggests the close relationship between *B. afzelii* migration and its rodent hosts. However, further studies into other small mammal host species of *B. afzelii* may aid in better understanding its ability to migrate.

5.3.5 Temporal analysis of Latvian sites

Qiu and colleagues (1997) investigated *B. burgdorferi* populations on Long Island and Shelter Island in the US over a three year period and observed little variation in the *ospA* and *ospC* types over space and time. This was similar to my findings for *B. garinii* as there was no differentiation of populations spatially and little evidence was found for temporal variation of clonal complex groups from Latvia. They suggested that these temporal and spatial data together show evidence of balancing selection in the *B. burgdorferi* populations maintaining allele frequencies over space and time (Qiu *et al.*, 1997). However, I maintain that my findings suggest host migration causes the homogeneity of *B. garinii* STs over space as I am able to contrast my *B. garinii* results with results from the rodent related strain, *B. afzelii* as discussed in section 5.3.1.

In Chapter 3 a trend of decreasing *B. afzelii* infections was observed at two of the three Latvian sites. Here it has been observed that the majority of strains lost between 2002 and 2006 were singleton STs that were not part of a common clonal complex group. These data suggest that this decline is having a greater impact on the singleton STs as opposed to the more common clonal complex groups. However, without further ecological data about the Latvian sites it is not possible to suggest the cause of the decline.

Chapter 6: Evolution of the LB group of spirochaetes

6.1 Introduction

Phylogenetic analyses of molecular data have advanced greatly over the past decade (Hall, 2008). Computation speed and more efficient algorithms have made more powerful tree building methods accessible to biologists. For example, Bayesian statistical approaches implemented in the program MrBayes, provide a rapid probabilistic method for phylogenetic inferences which produces equivalent or, in some cases, more accurate tree reconstruction than maximum likelihood (ML) (Hall, 2005, Ogden and Rosenberg, 2006). MrBayes searches for the most probable tree given the data while, ML reconstructs trees according to their fit to an evolutionary model. Recently, ML has become faster and easier to use due to localised tree improvement methods such as NNI and SPR (Hordijk and Gascuel, 2005) and the development of approximate likelihood ratio tests replacing time consuming bootstrap procedures (Anisimova and Gascuel, 2006, Guindon and Gascuel, 2003).

As mentioned earlier, the Lyme Borreliosis (LB) group comprises 17 named species. The geographic distribution and host specificity varies considerably among the members of the LB group of spirochetes (Table 6.1). There are also several putative species that are yet to be defined (Postic *et al.*, 2007). Two species, *B. afzelii* and *B. garinii* are regularly found across both Europe and Asia and two species occur in Europe and North America (*B. burgdorferi* and *B. bissettii*) (Masuzawa, 2004). The distribution range of the remaining species is more localised to regions within continents. For example, *B. japonica*, *B. turdi* and *B. tanukii* are localised to Japan (Masuzawa, 2004) while *B. lusitaniae* seems to be localised to sites around the Mediterranean Sea (Grego *et al.*, 2007). Originally only two species, *B. burgdorferi* and *B. bissettii*, were thought to be found in North America with *B. burgdorferi* the predominant species. However, recently four more species have been identified which are maintained by enzootic transmission cycles and appear to be more localised to regions within the United States (Postic *et al.*, 2007, Rudenko *et al.*, 2009a, Rudenko *et al.*, 2009b).

There is still debate about the nature of bacterial species and need for a theory-based prokaryotic species concept. As stated previously the *Borrelia* species were defined by DNA-DNA hybridisation which is based on the arbitrary value of 70 % association (Baranton *et al.*, 1992). However, in Chapter 4, I have shown that LB spirochetes form distinct entities. LB spirochaetes are comprised of distinctive ecotypes defined by their spectrum of vertebrate hosts, and the genetic data also supports the ecotype definition.

Using the MLSA scheme I have shown that strains believed to be of the same host specialism form tight clusters at the end of long branches. Here I have selected one or two representatives (where possible the type strain) from each ecotype to include in the evolutionary studies.

Table 6.1 A list of putative and named species within the LB group spirochaetes and their host range and distribution.

Species	Species status	Distribution	Host range	References
B. afzelii	confirmed	Europe and Asia	rodents	(Canica *et al.*, 1993)
B. americana	proposed	United States	Birds	(Postic *et al.*, 2007, Rudenko *et al.*, 2009b)
B. andersonii	confirmed	United States	Birds	(Marconi *et al.*, 1995)
B. bavariensis/ B. garinii serotype 4	proposed	Europe	Rodents	(Margos *et al.*, 2009)
B. bissettii	confirmed	United States and Europe	Rodents	(Postic *et al.*, 1998)
B. burgdorferi	confirmed	North America and Europe	Rodents and birds	(Johnsen *et al.*, 1984)
B. californiensis	confirmed	United States	unknown	(Postic *et al.*, 2007)
B. carolinensis	confirmed	United States	Rodents	(Rudenko *et al.*, 2009a)
B. garinii	confirmed	Predominantly Europe and Asia	Birds	(Baranton *et al.*, 1992)
B. japonica	confirmed	Japan	Rodents	(Kawabata *et al.*, 1993)
B. lusitaniae	confirmed	Mediterranean basin	Lizards	(Le Fleche *et al.*, 1997)
B. sinica	confirmed	China	Rodents	(Masuzawa *et al.*, 2001)
B. spielmanii	confirmed	Europe	Dor mouse	(Richter *et al.*, 2006)
B. tanukii	confirmed	Japan	Unknown	(Fukunaga *et al.*, 1996b)
B. turdi	confirmed	Japan	Birds	(Fukunaga *et al.*, 1996b)
B. valaisiana	confirmed	Europe and Asia	Birds	(Wang *et al.*, 1997)
B. yangtze	proposed	China	Rodents	(Chu *et al.*, 2008)
Genomospecies2	proposed	United States	Unknown	(Schwan *et al.*, 1993)
Strain 25015	putative	United States	Rodents	(Anderson *et al.*, 1988)

There is limited information available relating to the global evolution and relatedness of LB group species. Much of the previous work that included phylogenies of LB group spirochaetes was concerned with the definition of new species (Le Fleche *et al.*, 1997, Wang *et al.*, 1997). However, there has been debate over whether the common ancestor of *B. burgdorferi* evolved in the US or in Eurasia (Marti Ras *et al.*, 1997). Foretz and colleagues (Foretz *et al.*, 1997) used pulse-field gel electropheresis and arbitrarily primed PCR methods to compare the phylogenetic relatedness for *B. burgdorferi* strains. They found American strains to be more heterogeneous than European strains and concluded that *B. burgdorferi* most likely originated in North America and has since migrated to Europe. This conclusion was supported by Ras and colleagues using *ospC* as a

marker (Marti Ras *et al.*, 1997). However, Margos and colleagues (2008) used MLSA to create a phylogeny and included a group of highly divergent *B. burgdorferi* strains found in Switzerland, previously regarded as 'borderline' *B. burgdorferi* (Postic *et al.*, 2007). Using housekeeping genes, this group branched off deeply in phylogenetic trees from the other *B. burgdorferi* strains and it was suggested that this group represents an early European *B. burgdorferi* population. This observation, in combination with the higher species diversity in Europe, suggested that it is more likely that *B. burgdorferi* evolved in Europe and then migrated to the US. However, recently Rudenko and colleagues (Rudenko *et al.*, 2009b) created a highly inclusive phylogenetic tree of 16 of the LB group species and noted that all species found within North America, including many that have been recently defined, form a monophyletic group. This supports earlier suggestions by Postic and colleagues that *B. burgdorferi* originated in North America and was subsequently transmitted to Europe (Postic *et al.*, 1998).

Apart from the evolutionary history of *B. burgdorferi,* little has been reported in the literature about the evolutionary history of the LB group as a whole or the relatedness of the Eurasian strains. The choice of loci used in many previous studies did not allow the trees to be rooted because the genes do not exist outside of the LB group of spirochaetes. Here I utilise recently developed phylogenetic methods to ascertain the likely evolutionary history of the LB group species with consideration to their geographic spread as well as the evolution of host specialisation.

6.2 Results

The eight housekeeping genes were concatenated and used for analysis for 18 putative and confirmed LB group species (Appendix 8). In addition to these, several highly divergent strains from some species were also included in the analyses. These were *B. lusitaniae* PoHL1, *B. burgdorferi* Z41493, and *B. garinii* NT29. The only LB species missing from the dataset was *B. sinica*. The three relapsing fever spirochaetes, *B. hermsii*, *B. turicatae* and *B. duttonii*, were use as outgroup species for all trees and the concatenated sequence data for taxa not present on the MLST database (http://*Borrelia*.mlst.net) can be found in Appendix 8. Phylogenies were created using neighbour-joining (NJ), PhyML and MrBayes tree building methods as described in section 2.10.

While the different trees varied in overall topology, several species consistently grouped together (Figures 6.1 to 6.3). *B. tanukii*, *B. valaisiana* and *B. yangtze* formed a group in all three trees, which, for simplicity, will be referred to as the *B. valaisiana* group. *B. afzelii*, *B. spielmanii* were also always found adjacent. Furthermore, species found in

North America consistently formed a monophyletic clade with similar topologies within the clade in all three trees. *B. garinii*, *B. bavariensis*, *B. turdi* clustered together in both heuristic tree building methods but not in the NJ tree. *B. lusitaniae* and *B. japonica* were the only species that did not consistently group with any other species.

Each tree suggested a different species formed the most ancestral divide in the LB group. The NJ tree suggested that the species complex primarily divided between those species found in the US and those found in Eurasia (Figure 6.1). The Eurasian species appear to have under gone rapid speciation events creating the common ancestors of the species groups discussed above. The PhyML tree suggested that the primary branch from the LB group of species was the common ancestor of the *B. valaisiana* group clade albeit with low branch support. Apart from that, the tree supports a similar speciation pattern to that described for the NJ tree (Figure 6.2). The Bayesian phylogeny suggested that the primary species to branch off is *B. japonica* then the remaining Eurasian species groups branched off in quick succession with the final divide between the US strain common ancestor and the *B. valaisiana* group common ancestor (Figure 6.3).

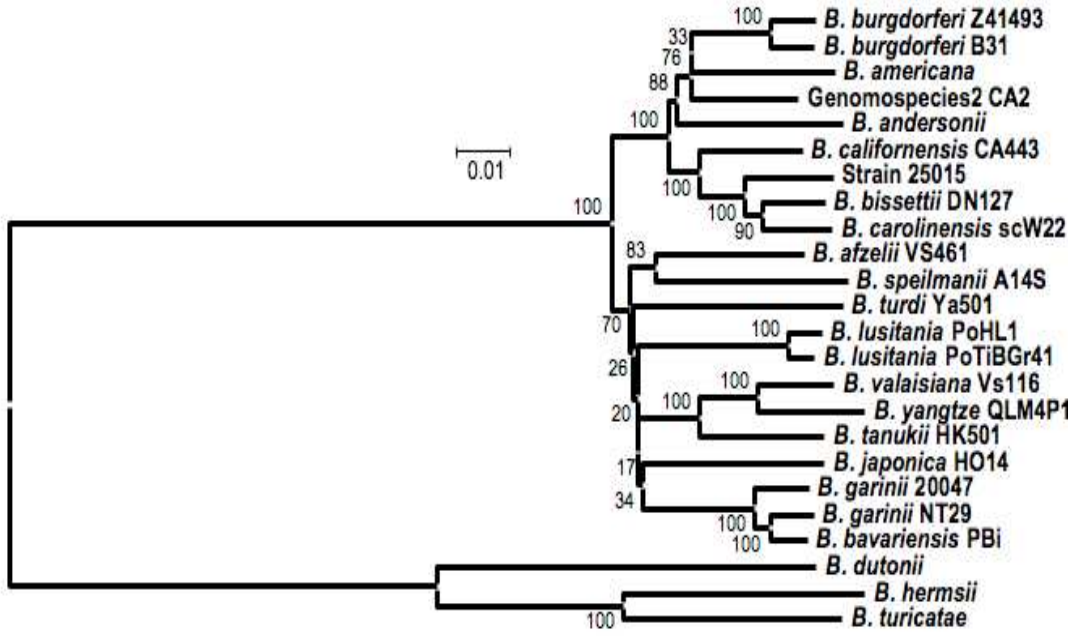


Figure 6.1 Neighbouring-joining tree showing LB groups species. The tree is rooted using *B. duttonii*, *B. hermsii* and *B. turicatae*. The scale bar shows 1 % divergence. Branch confidence calculated by bootstrap method.

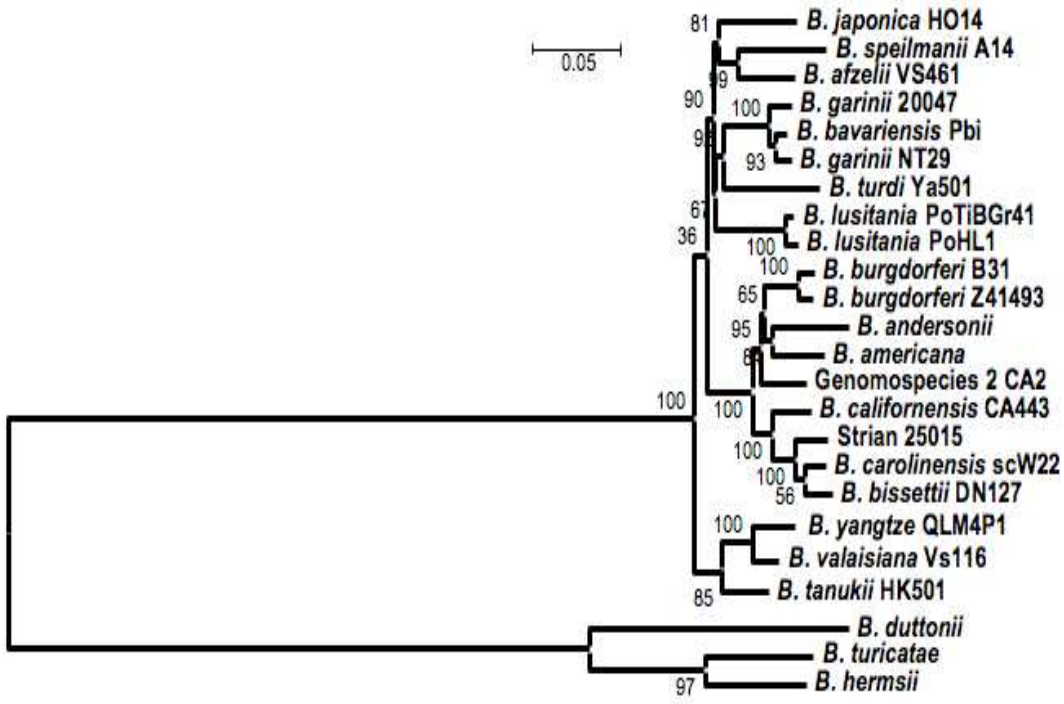


Figure 6.2 PhyML tree showing LB groups species.The tree is rooted using *B. duttonii*, *B. hermsii* and *B. turicatae*. The scale bar shows 10 % divergence. Branch confidence values calculated using aLRT.

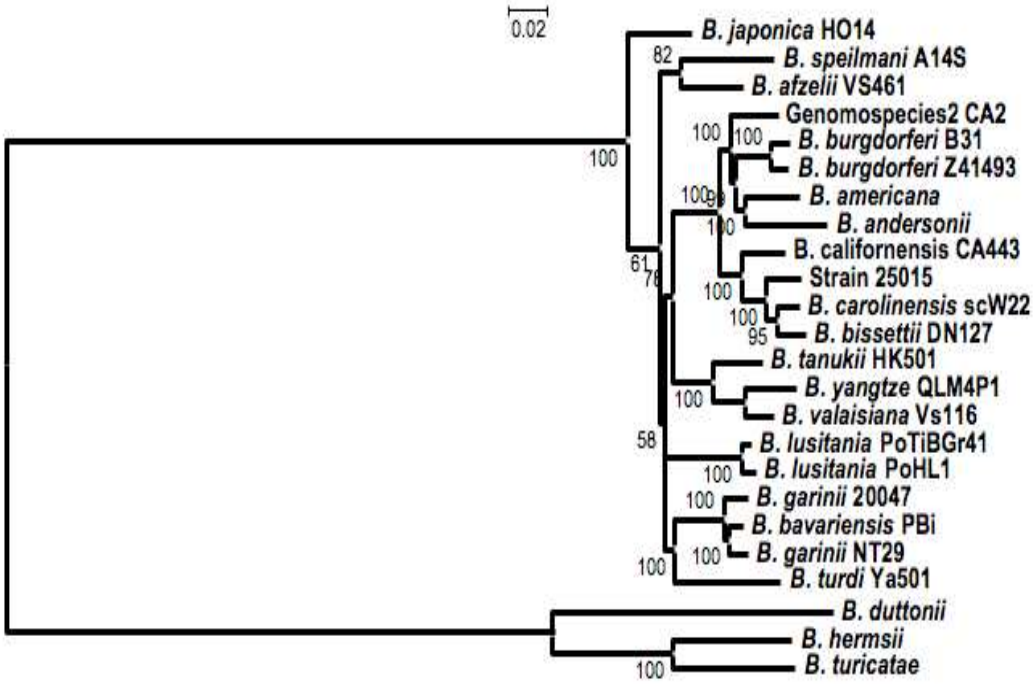


Figure 6.3 Bayesian phylogeny showing LB groups species. The tree is rooted using *B. duttonii*, *B. hermsii* and *B. turicatae*. The scale bar shows 2 % divergence. Branch confidence values calculated using posterior probability.

The Robinson and Fould distance (RFd) between the two heuristic topologies was calculated as described in section 2.11.5. This method determines the percentage of shared bi-partitions between two trees. It was found that 83 % of bipartitions are shared between

the Bayesian and PhyML topologies illustrating that they are, in fact, fairly similar topologies and as stated above, it appears that nearly all the incongruence occurs at the deep internal nodes. All the trees show extremely short branching at many of the deep internal nodes suggesting a rapid radiation of species may have occurred early in the evolutionary history of the LB group. As stated in chapter 4, these short branches mean that there are few polymorphic sites in the sequence data to define these short branches or alternatively may be suggestive of incomplete lineage sorting (Avise and Robinson, 2008, Maddison, 1997). In light of these observations two alternative theories were considered to aid in the identification of the true tree for the LB group spirochaetes. First I will discuss the potential occurrence of incomplete lineage sorting in my data set and secondly I will investigate whether the inclusion of additional loci improves the phylogenies.

6.2.2 Incomplete Lineage Sorting

6.2.2.1 Individual gene phylogenies

Phylogenies were created for each of the housekeeping genes individually using both PhyML and MrBayes (section 2.10). On visual inspection of the topologies and through the use of RFd it was observed that there was wide variation of topologies created by the different genes (Table 6.2, Appendix 5). Using the PhyML inference method the US species group and *B. valaisiana* species group were each found to be most ancestral branch in three out of the eight gene trees. The majority rule consensus tree of the eight individual housekeeping gene trees, created using PHYLIP (section 2.11.5), suggested the *B. valaisiana* species group was the most ancestral branch (Figure 6.4). This was in agreement with the PhyML tree created using the concatenated gene sequence (Figure 6.2, Table 6.2). The equivalent majority rule consensus tree of the Bayesian gene trees was also fairly congruent with the Bayesian inference of the concatenated genes (Figures 6.3 and 6.5, Table 6.2). However, the individual gene trees created using MrBayes frequently contained polytomies at the deeper branches and the individual gene trees rarely agreed on the species that formed the most ancestral branch.

When considering the congruence between the two inference methods it was observed that four genes (*nifS*, *pepX*, *pyrG* and *recG*) showed high congruence between the two phylogenetic inference methods, sharing between 86 and 90 % of bipartitions (Table 6.2). However, when compared visually there were still differences between these individual housekeeping gene trees concerning the most ancestral branch (Appendix 5). The remaining genes showed medium (*clpA*, *rplB* and *uvrA*) or low (*clpX*) levels of congruence between the two methods and the Bayesian inferences frequently contained

polytomies suggesting many relationships could not be resolved using the data available. These results may suggest that there are too few polymorphic sites within the individual gene sequence data to fully resolve the trees (Table 6.3). However, the RFd in Table 6.2 did reveal that the *clpX* topology appeared to differ substantially from any other using the PhyML inference method and to a high degree using the Bayesian inference method which may indicate it has a different evolutionary history to the other genes.

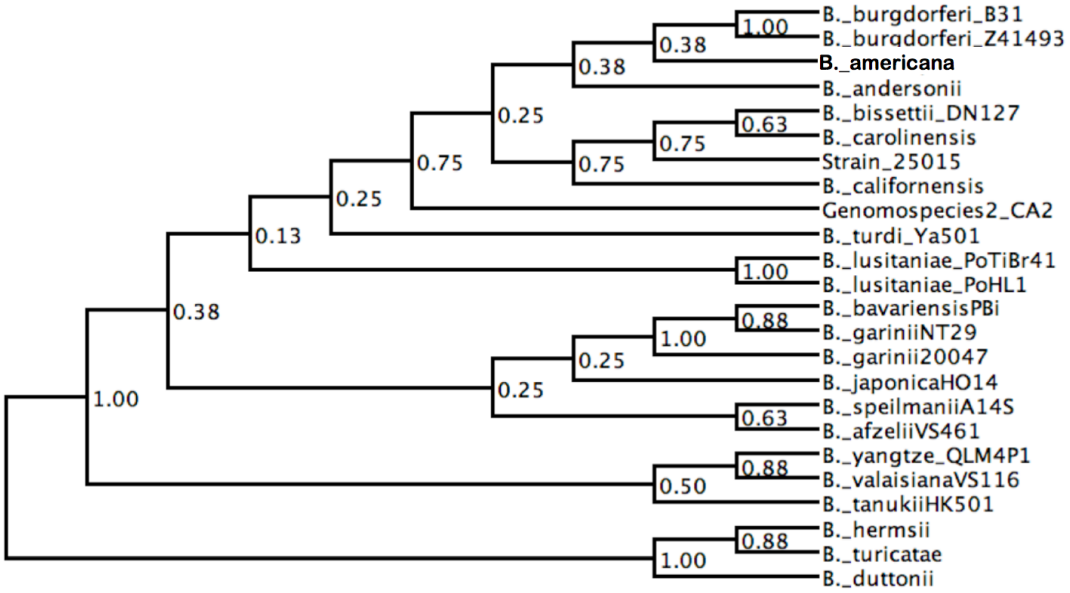


Figure 6.4 Phylip consensus tree of the eight individual PhyML housekeeping gene topologies. The branch values represent the proportion of individual trees that contained that node.

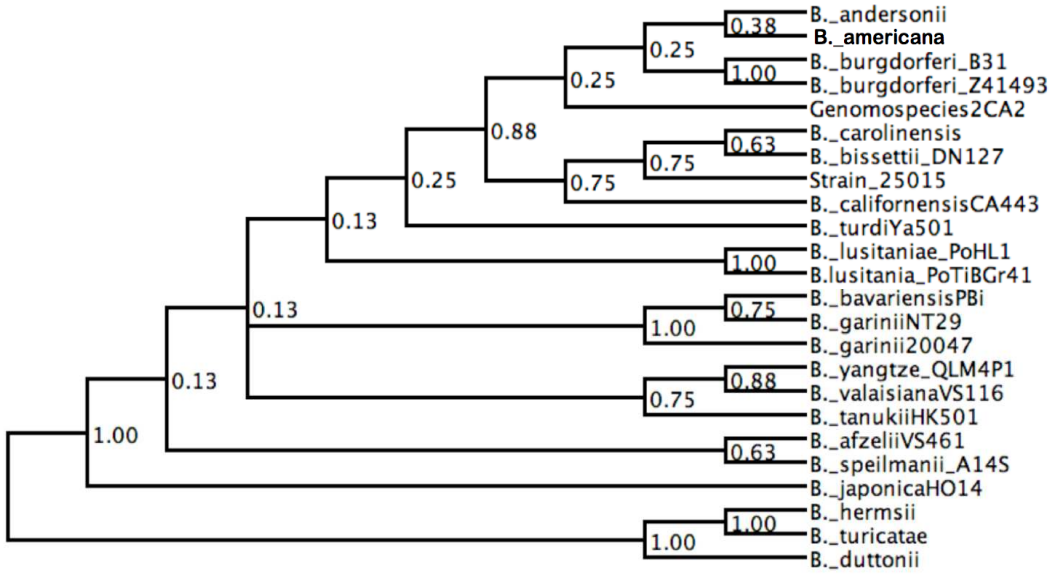


Figure 6.5 Phylip consensus tree of the eight individual Bayesian housekeeping gene topologies. The branch values represent the proportion of individual trees that contained that node.

Table 6.2. RFd values for individual and concatenated topologies are shown. These were calculated using either Bayesian or PhyML inference methods. Values are coloured in reference to the congruence of the two topologies; 20-39 % of bipartitions in common are coloured light grey through to 80-99 % matcing bipartitions are coloured black. Topologies that are 100 % identical are left white.

	Percentage of matching bipartitions																									
PhyML	clpA	|	100																							
	clpX	|	33	100																						
	nifS	|	62	38	100																					
	pepX	|	48	38	48	100																				
	pyrG	|	38	33	52	48	100																			
	recG	|	67	29	48	48	33	100																		
	rplB	|	67	29	62	43	43	52	100																	
	uvrA	|	52	29	52	48	38	48	48	100																
	phyML	|	67	33	62	62	52	71	71	57	**100**															
	Bayesian	|	60	36	64	64	55	60	64	60	**83**	**100**														
	Phyml majority rule	|	76	33	81	52	52	57	81	57	**71**	**69**	**100**													
	Bayes majority rule	|	60	36	69	60	50	60	74	60	**79**	**90**	**79**	**100**												
	No pyr phyML	|	71	38	57	62	48	76	62	57	**90**	**74**	**67**	**69**	**100**											
	Nopyr Bayes	|	60	36	64	64	55	60	64	60	**83**	**95**	**69**	**86**	**74**	**100**										
	No clpX phyML	|	67	33	62	62	52	71	71	57	**100**	**83**	**71**	**79**	**90**	**83**	**100**									
	No clpX bayes	|	52	33	57	67	57	57	57	52	**81**	**93**	**62**	**83**	**71**	**88**	**81**	**100**								
Baysian	clpA	|	67	33	62	57	43	62	57	57	71	74	67	69	76	79	71	67	100							
	clpX	|	62	48	52	48	48	57	67	52	57	60	57	60	57	60	57	52	57	100						
	nifS	|	67	48	86	57	57	57	67	62	71	79	81	83	67	79	71	71	71	62	100					
	pepX	|	50	40	50	93	50	50	50	50	69	71	55	67	64	71	69	74	60	50	60	100				
	pyrG	|	52	48	67	62	86	48	57	52	67	69	67	64	62	69	67	71	57	62	71	64	100			
	recG	|	71	38	57	57	43	90	62	57	81	69	67	69	81	69	81	67	67	67	67	60	57	100		
	rplB	|	60	40	79	55	55	55	74	60	69	76	79	86	64	71	69	69	64	60	79	62	69	64	100	
	uvrA	|	69	50	69	74	64	64	64	74	74	76	74	76	74	76	74	79	74	69	79	76	79	74	76	100

Table 6.3 Shows the length of the alignment, number and percentage of variable sites, mean p-distance and dN/dS ratio for all taxa included in the phyologeneitic analyses. For the individual housekeeping genes and the concatenated genes this refers to all 21 ingroup taxa while, for *recA*, *flaB*, and *hbb* this refers to the 13 ingroup taxa of the extra gene set.

Data	No. sites	No. variable sites	Percentage variable sites	p-distance	dN/dS
clpA	579	208	35.9	0.091	0.284
clpX	624	134	21.5	0.055	0.044
nifS	564	137	24.3	0.065	0.131
pepX	570	172	30.2	0.076	0.195
pyrG	606	173	28.5	0.074	0.118
recG	651	168	25.8	0.072	0.132
rplB	624	138	22.1	0.053	0.098
uvrA	570	148	26.0	0.068	0.065
Concatenated	4788	1275	26.6	0.069	0.382
recA	160	55	34.4	0.091	0.173
flaB	196	46	23.5	0.063	0.388
hbb	327	60	18.3	0.050	0.268
Extra gene concat	5470	1186	21.7	0.066	0.350

6.2.2.2 Confidence sets for gene trees

The concatenated and individual gene alignments were compared to all their respective topologies using TREE-PUZZLE 5.3 as discussed in section 2.11.4. The test uses four mathematical models to assess the likelihood that an alignment would accept a specific topology. There was great variation between the models as to whether a particular alignment accepted a specific topology (Table 6.4). However, the consensus between the methods was that *pyrG* alignment consistently rejected all topologies (other than its own) and furthermore the *pyrG* topology was rejected by most alignments as well (Table 6.4). This suggests that this gene may have had a different evolutionary history and may not be compatible with the other genes included in this study. Interestingly, while the *clpX* topology was rejected by most alignments, the *clpX* alignment did not reject many other gene topologies.

The analyses, however, did not clearly favour either of the concatenated gene trees, Bayesian or PhyML, over the other. Three of the individual gene alignments were considered to fit better with the Bayesian concatenated gene topology, three were considered to fit better with the PhyML concatenated gene topology and two (*nifS* and *pyrG*) most often rejected both topologies using the different statistical tests (Appendix 6). When using the concatenated housekeeping gene alignment TREE-PUZZLE suggested that, while both concatenated gene topologies were accepted by the alignment, the PhyML tree was identified as the better topology for the data (Table 6.4). However, the Bayesian phylogeny was accepted by more individual gene alignments and mathematical methods than the PhyML tree (Table 6.4).

Table 6.4. Confidence sets showing whether different gene alignments accept the tree topologies of the PhyML inferences of the individual housekeeping genes and the PhyML and Bayesian inferences of the concatenated housekeeping genes. Topologies accepted by the alignment are represented by a plus sign in a shaded square. Whether the topologies were accepted by the alignment was tested using four mathematical tests and these are labelled 1 to 4. The numbers represent the following tests; 1) one-sided Kishino-Hasegawa test, 2) Shimodaira-Hasegawa test, 3) expected likelihood weights and 4) two-sided Kishino-Hasegawa test which also indicates the topology that best fits the alignment (B). The concatenated phylogenetic inferences for both PhyML and Bayesian were compared to the gene alignments.

	Alignment																															
Tree	***clpA***				***clpX***				***nifS***				***pepX***				***pyrG***				***recG***				***rplB***				***uvrA***				**Concat.**			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
ClpA	+	+	+	B	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	+	-	+	-	+	+	+	+	+
clpX	-	-	-	-	+	+	+	B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
nifS	-	+	-	-	+	+	+	+	+	+	+	B	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+	+	+	+	+	-	+	-	-
pepX	-	+	-	+	+	+	-	+	-	+	-	-	+	+	+	B	-	+	-	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
pyrG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
recG	+	+	+	+	+	+	-	+	-	+	-	-	-	+	-	-	-	-	-	-	+	+	+	B	-	+	-	-	+	+	-	+	-	+	-	-
rplB	+	+	-	+	+	+	+	+	-	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	+	+	B	+	+	-	+	-	+	-	-
uvrA	-	+	-	-	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-	-	+	-	+	-	+	-	-	+	+	+	B	-	+	-	-
phyML	+	+	+	+	+	+	-	+	-	+	-	-	+	+	-	+	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	B
Bayesian	+	+	+	+	+	+	-	+	-	+	-	+	+	+	+	+	-	-	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+

6.2.2.3 Exclusion of potentially incompatible genes

The confidence sets comparing alignments and tree topologies as well as the RFd results suggested that *clpX* and *pyrG* may not be compatible with each other or the majority of the other housekeeping genes so they were alternately removed to investigate the influence each gene had on the concatenated topology. PhyML and Bayesian topologies were constructed excluding each gene (as stated in section 2.10) and compared to the original inferences based on topology and branch support values.

The removal of *pyrG* and *clpX* from the alignment had limited, or in one case, no effect of the tree topologies or on branch support values using either the Bayesian and PhyML inference methods (Table 6.2, Appendix 7). The trees reconstructed after excluding *pyrG* differed from the Bayesian and PhyML trees that used the complete gene alignment by only one and two bipartitions. Similarly, the no *clpX* trees differed by 2 bipartitions, in the case of the PhyML method, and was identical in topology using the Bayesian inference method (Table 6.2). In all Bayesian inferences *B. japonica* was consistently the most ancestral branch and in the PhyML inferences the *B. valaisiana* group always formed the most ancestral branch.

6.2.3 Inclusion of Additional Loci

The Genbank database was searched for potential housekeeping genes available for the majority of *Borrelia* species to investigate whether the inclusion of additional genetic

information aids in the identification of the true evolutionary history of the LB group spirochaetes. Three housekeeping genes were selected on the basis that they were available in the majority of LB group species, these were *flaB*, *hbb* and *recA*. In terms of dN/dS ratio and mean pairwise distance they were all found to be comparable to the MLSA housekeeping genes albeit shorter in length (Table 6.3). Sixteen taxa of the original 24 were available for all three genes for the same strain, these will be referred to as the 'extra gene' taxa set. The genes were not available for the *B. afzelii* type strain, VS461, which has been used in the study until this point, but all the eight MLSA housekeeping genes and all three extra genes were available were available in *B. afzelii* strain PKo so this was used instead. The same situation occurred for both *B. lusitaniae* strains which were removed and replaced with *B. lusitaniae* PotiB2.

PhyML and Bayesian inferences were created (section 2.10) using the concatenated alignment of the MLSA scheme housekeeping genes and the three extra housekeeping genes. These topologies were compared to MLSA gene topologies where i) the original MLSA gene tree topologies (Figures 6.2 and 6.3) were pruned for the 16 taxa for which all genes were available (so taxa not found in the extra gene taxa set were pruned out to allow for direct comparison with the extra gene tree) and, (ii) where the concatenated MLSA gene trees were remade including only the extra gene taxa set.

The three extra genes altered the topology of the Bayesian inference so that *B. valaisiana* formed the most ancestral branch in the phylogeny (Figure 6.7). In the PhyML inference *B. valaisiana* remained the most ancestral branch so for the first time both the PhyML and Bayesian inference supported the same ancestral branch (Figure 6.6). However, there was no improvement in branch support values. Furthermore, using TREE-PUZZLE, it was shown that the Bayesian phylogeny was not supported by many of the individual gene alignments and the concatenated alignment of the MLSA genes did not support the topology under any of the statistical methods (Table 6.5). Overall, all the three Bayesian topologies were considered less compatible than the PhyML alternatives when compared to the gene alignments.

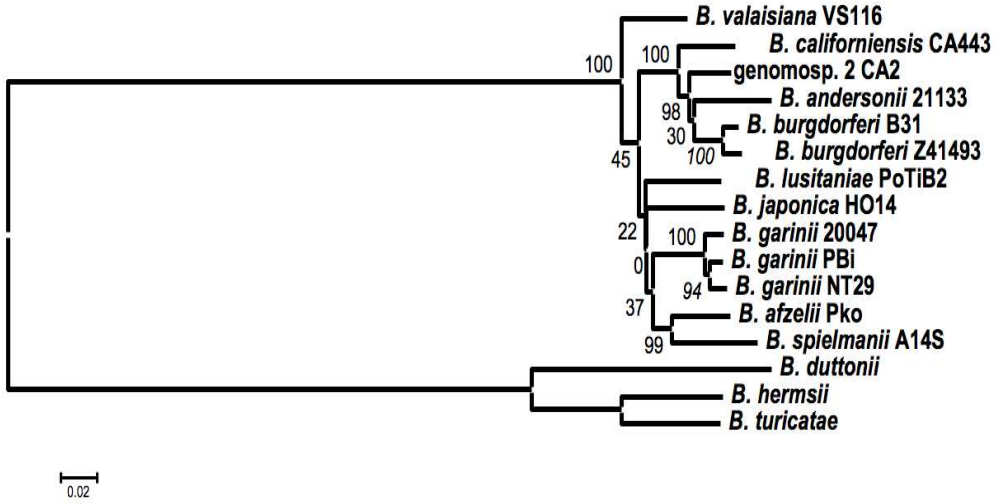


Figure 6.6 PhyML phylogenetic inference of LB group of species using the MLSA housekeeping genes and the three extra gene regions of *recA*, *hbb* and *flaB*. Three relapsing fever spirochaetes, *B. duttonii*, *B. hermsii* and *B. turicatae* were used as outgroup species and the scale bar shows 2 % divergence.

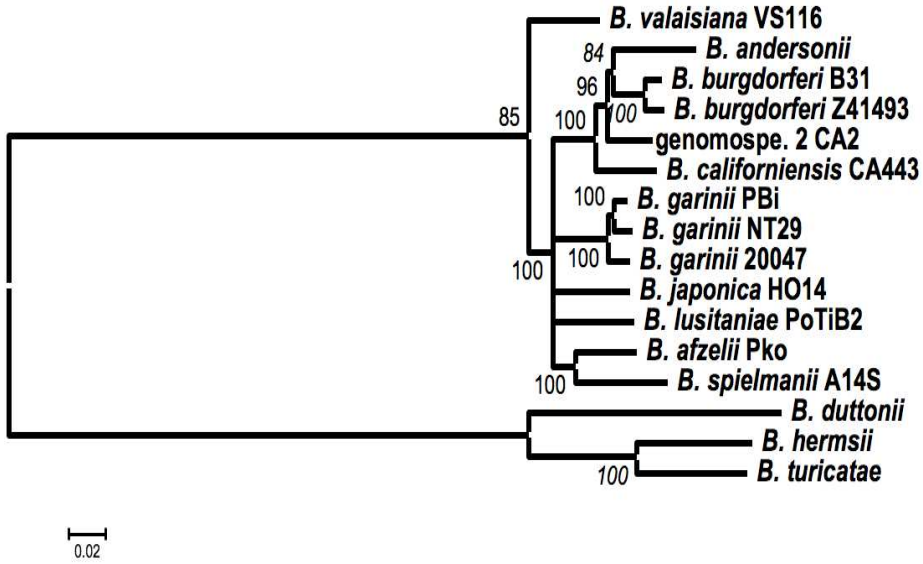


Figure 6.7 Bayesian phylogenetic inference of LB group species using the MLSA housekeeping genes and the three extra gene regions of *recA*, *hbb* and *flaB*. Three relapsing fever spirochaetes, *B. duttonii*, *B. hermsii* and *B. turicatae* were used as outgroup species and the scale bar shows 2 % divergence.

Table 6.5. Confidence sets showing whether different gene alignments accept the PhyML and Bayesian tree topologies which were inferred using the MLSA genes alone, including the three extra genes or pruned versions of the original trees. Topologies accepted by the alignment are represented by a plus sign in a shaded square. The numbers one to four represent the following mathematical tests; 1) one-sided Kishino-Hasegawa test, 2) Shimodaira-Hasegawa test, 3) computes the expected likelihood weights and 4) two-sided Kishino-Hasegawa test which also indicates the topology that best fits the alignment (B).

	Alignments																																			
Trees	***clpA***				***clpX***				***nifS***				***pepX***				***pyrB***				***recG***				***rplB***				***uvrA***				***flaB***				***hbb***				***recA***				**concat**			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
phyml extra genes	**+**	**+**	**+**	**B**	+	+	+	+	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**B**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
phyML prune	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**B**	+	+	+	+	+	+	+	**B**	+	+	+	**B**	+	+	+	+	+	+	+	**B**	+	+	+	+
phyML remake	+	+	+	+	+	+	+	**B**	+	+	+	**B**	+	+	+	+	+	+	+	**B**	+	+	+	+	+	+	+	**B**	+	+	+	+	+	+	+	**B**	+	+	+	+	+	+	+	+	+	+	+	**B**
Baysian extra genes	**-**	**-**	**-**	**-**	+	+	-	+	**+**	**+**	**-**	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**-**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	+	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
Bayesian pruned	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	**B**	+	+	-	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	-	+
Bayesian remake	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	-	-	-	+	-	-

RFd was used to compare the tree topologies of the six trees. It was observed that there was substantial variation between the topologies that were obtained via the three different methods. Under the PhyML inference method the pruned MLSA tree only matched the remade MLSA tree 62 % of bipartitions suggesting the removal of the eight taxa caused the topology to change (Table 6.6). However, there was good congruence between the pruned PhyML tree and the extra gene tree. A similar pattern occurred using the Bayesian inference method where there was less congruence between the two MLSA trees than between the MLSA trees and the extra gene tree (Table 6.6). While the Bayesian tree had altered in such a way that *B. valaisiana* now formed the most basal branch, there was only a minimal increase in congruence between the PhyML and Bayesian topologies for the extra gene trees (88 % matching bipartitions, Table 6.6) compared to the original MLSA topologies (83 % matching bipartitions, Table 6.2).

Table 6.6. Shows RFd comparing the extra gene tree topologies (in bold) with the MLSA gene topologies when either the additional taxa were pruned out or the tree was remade without them. The boxed regions highlight the RFds comparing trees created using the same inference method.

RFd percentage of matching partitions						
Phyml extra	**100**					
Phyml pruned	**85**	**100**				
Phyml remake	**69**	**62**	**100**			
Bayes extra	88	88	73	**100**		
Bayes pruned	73	73	58	**85**	**100**	
Bayesian remake	65	65	88	**77**	**69**	**100**

The limited improvement in branch support values or confidence in a certain topology may be due to the fact that while there was an increase in genetic information used for those taxa it was available for, overall there was a net decrease in genetic information in the alignment of approximately 27,000 bp due to the loss of eight taxa. Furthermore, quartet puzzling analysis using TREE-PUZZLE (section 2.11.4) revealed that both the MLSA alignment (with more taxa) and the extra gene alignment (with more genes) were able to resolve a similar number of quartets (95.9% and 96.2 % respectively). This suggests that while both alignments are likely to be good at identifying tree topology, neither is likely to have more power or ability to do this.

6.3 Discussion

6.3.1 Evolution of the LB Species Group

This study has produced the most complete LB species tree inferring the global evolution of the LB group of spirochaetes. Prior to this study Masuzawa and colleagues (2004) have created the most complete rooted phylogeny of the group. It included 11 different ingroup species taxa and was constructed using a NJ algorithm. Similar to the NJ tree contructed by this study, it also suggested the most ancestral branch occurred between the US and Eurasian species. The phylogeny was entirely based on a fragment of the *flaB* gene and so was based on limited genetic information. The tree published by Masuzawa did include the species *B. sinica* that is missing from my study. Its position in the tree suggests that the species is highly distinct from many of the LB species and places it nearest to *B. garinii* and *B. japonica* (Masuzawa *et al.*, 2004). Interestingly, the unrooted phylogeny by Rudenko and colleagues (2009) places *B. sinica* adjacent to *B. japonica* but not adjacent to *B. garinii*. This suggests that it may be worthwhile endeavouring to add *B. sinica* to the MLSA data set in future studies as it may provide useful extra information to the phylogenies.

Although there was conflicting information on the deeper branching of the LB group of species, there were many areas of agreement between different phylogenetic inferences about its evolution, these factors have been summarised in Figure 6.8. Strains found in North America formed a monophyletic group and suggest that early in the evolutionary history of the LB group there was a split between the ancestral species found in Eurasia and those in North America. Interestingly, two North American species, *B. burgdorferi* and *B. bissettii*, are found in Europe as well but they do not appear to be ancestral species of the New world group. This suggests that these species have returned to Europe substantially later in the evolution of the New World group by unknown means. This phylogenetic divide between species found in Eurasia and North America has been noted previously (Postic *et al.*, 2007, Rudenko *et al.*, 2009b) but they were unable to make definitive claims about the direction of the migration. The majority of phylogenies, including both majority rule phylogenies (Figures 6.4 and 6.5) suggest the most ancestral branching occurs between species found in Eastern Asia (although these are different speices in the PhyML and Bayesian inferences) suggesting a Eurasian origin for the species group.

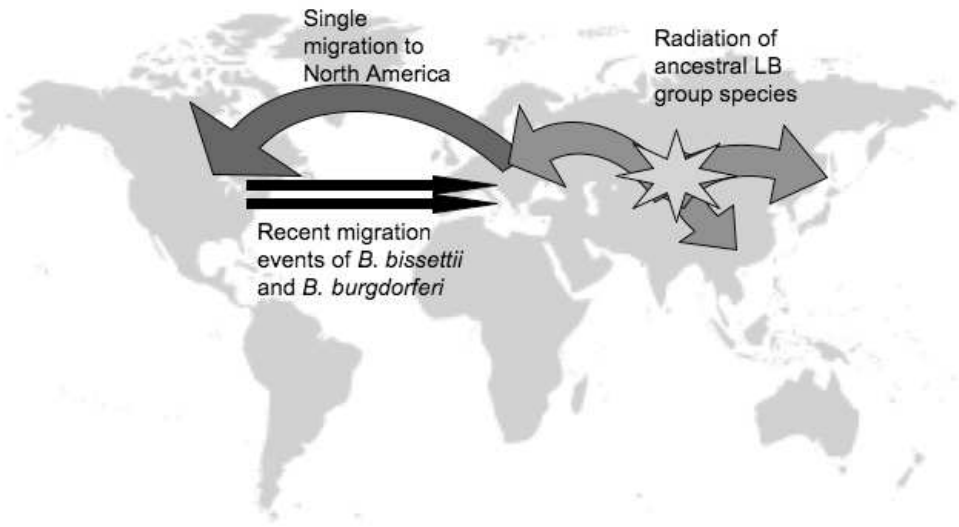


Figure 6.8. My theory of the geographic evolution of the LB group of species. The theory is based on areas of consensus between the Bayesian and PhyML phylogenies. Light grey symbols indicate events in the distant evolutionary past while darker colours represent events that happened in more recent evolutionary time.

Interestingly, the suggestion of a Eurasian origin echos the currently favoured hypotheses for the evolution of the *Ixodes persulcatus* species complex. The *I. persulcatus* species complex includes the vast majority of known competent vectors for the LB species group and all species in the *I. persulcatus* group that have been tested have been found to be vector competent for the LB group of species (Eisen and Lane, 2002b). Filippova (Filippova, 1973) suggested that the *I. persulcatus* species complex evolved in either Eurasia and that ticks then migrated via the Bering Straight to North America or that the complex evolved in Laurasia and were separated after the continental split and that adaptation and speciation then occurred independently.

Studies into the evolution of the *I. persulcatus* complex also suggest a rapid radiation of species possibly due to habitat changes or changes in the availability of mammalian hosts (Filippova, 1991, Xu *et al.*, 2003). A similar pattern was observed in the LB group phylogenies as rapid consecutive speciation events occurred within the Eurasian species group.

While these factors suggest a close evolutionary relationship between the tick complex and the LB group of species, there are apparent differences between their evolutionary histories (Humphrey *et al.*, 2010, Qiu *et al.*, 2002). Studies have noted that there is little coevolution between *I. scapularis* ticks and *B. burgdorferi* in the US. Furthermore, on a broader scale several studies have indicated that *I. persulcatus,* along with other East Asian tick species cluster with the western US tick species, *I. pacificus* (Figure 6.9, *(Xu et al., 2003).* The majority of US species found in the species complex, such as *I. minor* and *I. affinis,* cluster in a separate sister clade with the European tick

species, *I. ricinus*. *I. scapularis* appears to cluster separately from both of these sister clades (Xu *et al.*, 2003). This suggests that there were multiple introductions of *I. persulcatus* species complex ticks into the US. It has also been suggested that this may represent migration across the Bering Straights (Filippova, 1973, Fukunaga *et al.*, 2000, Xu *et al.*, 2003). However, here no such close link was been observed between Asian LB species and LB species found on the west coast of the US.

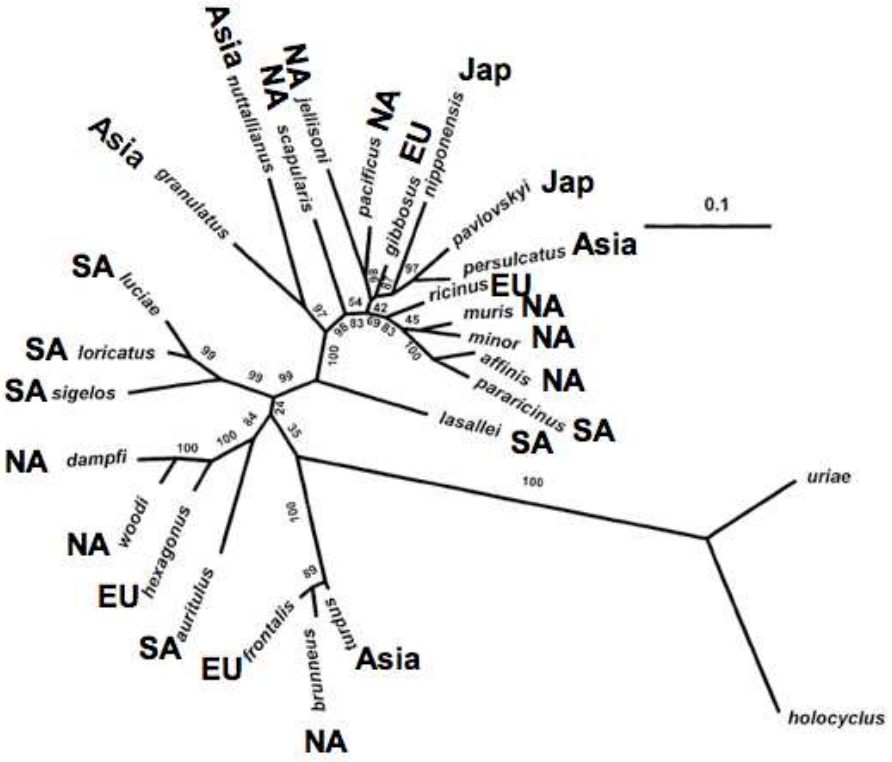


Figure 6.9 Bayesian phylogeny by Xu and colleagues (2003) showing species of the *Ixodes persulcatus* species complex. The global region the tick is usually found is indicated by the following codes; EU Europe, NA North America, SA South America, Jap Japan and Asia. *I. uriae* and *I. holocyclus* are not part of the species complex.

Interestingly, the US LB species, as a whole, do not appear to have speciated based on geographic isolation as species such as *B. californensis* and *B. carolinensis* occur in the same US sub-clade, but are found on opposite sides of the continent. The US group of LB species consistently subdivided into two clades, which appear to differ by host specialisation. The clade containing *B. carolinensis*, *B. bissettii*, *B. californensis* and the putative species that includes strain 25015 are all know to have rodent related hosts. However, the second clade contained *B. burgdorferi*, a generalist, *B. andersonii* and *B. americana*, both known to have bird related hosts, and Genomospecies 2 the host range of which remains unknown.

Thus while I have shown evidence suggesting a potential link between the evolution of the vector species complex and the LB species complex, they do not appear to be as tightly linked into coevolutionary relationships as other vector/pathogens pairs such as mites (for a review see (Proctor and Owens, 2000). This may be due to different evolutionary forces acting upon the ticks compared to the spirochaetes. For example it has been suggested that ticks must adapt to the environment they find themselves in, be this an animal burrow or woodland floor (Klompen *et al.*, 1996). Although the spirochaetes are reliant on the ticks as vectors, they must also adapt to survive within host species they are likely to come into contact with and this subtle difference may or may not cause coevolution to occur between the spirochaetes and the ticks. This may partially explain why in relatively small scale investigations in the US into the coevolution of *I. scapularis* and *B. burgdorferi* have found no evidence for it (Humphrey *et al.*, 2010, Qiu *et al.*, 2002). It seems likely that LB species that rely in nidicolous tick species, such as *B. bissettii*, will in future have much closer coevolutionary relationships than other *Borrelia* and their associated non nidicolous species as more tick and spirochaete have similar environmental selective pressures.

Whether the two US sub-clades of different host specialisms appear this way through coincidence is unknown but the same clustering of species of certain host specialisations was not observed in Eurasian strains. For example, potentially one of the most recent speciation events occurred between *B. garinii*, a bird related species and *B. bavariensis*, a rodent related species (Margos *et al.*, 2009). The evolutionary history of LB species in Eurasia suggests that adaptation to different hosts has repeatedly happened. For example, there is no single recent common ancestor for all bird associated or rodent associated species. This example, among others, suggests that host specialisation can change fairly readily.

6.3.2 Incomplete Lineage Sorting and the MLSA Gene Set

While the TREE-PUZZLE confidence set tests revealed that some gene loci (*pyrG* and *clpX*) may have contrasting evolutionary histories to the majority of the MLSA genes, they did not substantially influence the concatenated tree topology. This further demonstrates that the MLSA scheme is a robust method for investigating the LB group of spirochetes. Previously it has been shown that the MLSA genes have similar evolutionary rates (Loza-Reyes, unpublished) but this is the first time it has been possible to compare the species complex as a whole.

There was generally consensus between the various LB phylogenies and over 95% of quartets were resolved. It therefore seems promising that the addition of the three extra genes may further resolve the trees, but a greater number of taxa is still required.

pyrG is a candidate to test for the occurrence of incomplete lineage sorting due to the rapid speciation events described above and in Chapter 4 (section 4.4.1, Figure 4.6) as well as the TREE-PUZZLE confidence sets which suggested that the *pyrG* alignment was not compatable with the other gene topologies. I suggest that further studies such as coalescent simulations should be carried out to further investigate this possibility.

Chapter 7: General Discussion

7.1 Chapter Summary

- **Chapter 3**
  - Ticks appeared well established in woodland areas of Wiltshire and Avon and their densities were comparable with other areas of continental Europe.
  - The LB infection prevalence in ticks ranged greatly between sites in England.
  - Although there has been a lack of published data on LB infections in England, here I have reported for the first time the presence of *B. afzelii* in England.
- **Chapter 4**
  - In this chapter the successful development of a MLSA scheme for LB group of spirochaetes based on eight housekeeping genes is described.
  - The scheme revealed that the LB species form distinct ecotypes and that there is limited horizontal gene transfer within and between the LB group species.
- **Chapter 5**
  - Using the MLSA scheme it was revealed that differences in the level of geographic structure in populations of LB species was consistent with patterns of migration of their different vertebrate hosts.
  - Three LB species investigated, *B. afzelii*, *B. garinii* and *B. valaisiana* all show evidence of sudden population expansion in their past.
  - The rodent related species, *B. afzelii* showed evidence of fine scale structuring at sites within countries which may parallel with structuring observed in its rodent hosts.
- **Chapter 6**
  - The phylogenies of the LB group of species suggested a Eurasian origin for the group and that early in their evolutionary history new ancestral species evolved in quick succession.
  - Also early in the evolutionary history LB strains appear to have become isolated in North America and diverged independently from those in Eurasia.
  - In more recent time strains from two North American species appear to have migrated back to Europe.

7.2 Major Findings

The major objective of this thesis was to further the understanding of the migration and spread of LB group of species as well as to attempt to resolve the evolutionary geographic origin of both the group as a whole and the species it contains. Key to these issues is the availability of an appropriate genotyping tool which provides the resolution required for population studies within species but which is also applicable to all species within the LB group. Previously developed typing schemes are either based on single or multiple loci, several of which are exclusive to the LB group making it impossible to root trees. It has been suggested that the very conserved loci, such as *flaB* and 16S rRNA, contain too few polymorphic sites and are not sensitive enough for population studies within species (Bunikis *et al.*, 2004). For this reason, with Klaus Kurtenbach and Gabriele Margos, I have aided in the development of a novel multilocus sequence analysis (MLSA) scheme for the LB group of spirochaetes.

7.2.1 MLSA system

This scheme is a typical MLSA scheme based on regions of eight housekeeping genes (*clpA, clpX, nifS, pepX, pyrG, recG, rplB* and *uvrA*) located on the linear chromosome of the LB group of spirochaetes. The scheme has been revealed to have the epidemiological power to (i) establish phylogenetic relationships among the *Borrelia* populations (Margos *et al.*, 2008), (ii) delineate new species (Margos *et al.*, 2009), (iii) assign infections/isolates to species (Margos *et al.*, 2009, Vollmer *et al.*, 2010), (iv) capture genetic structure within species (Vitorino *et al.*, 2008, Vollmer *et al.*, 2010) and (v) identify intraspecific geographic structure of the bacteria (Vollmer *et al.*, 2010).

The scheme further benefits from the fact that it is possible to amplify *Borrelia* DNA directly from infected tick material. However, whilst this removes any bias that may be created through culturing as well as making the process much more rapid, mixed infections within questing ticks cannot be readily characterised through sequencing. Using the scheme I have shown that LB group species appear to have extremely low rates of intra-species horizontal gene transfer on their chromosome compared to other bacterial species (Vos and Didelot, 2009). In my work, no cases of interspecies chromosomal gene transfer were observed and each species formed a distinct closely related cluster at the end of long branches. The only evidence for interspecies gene transfer was apparent in *ospA* tree. This is a plasmid located gene and evidence of gene transfer has been observed previously at this locus as well as other plasmid located genes (Lin *et al.*, 2002, Wang *et*

al., 1999a). This suggests there may be a higher, but still relatively rare, rate of interspecies gene transfer on plasmids.

This is the first multilocus approach designed to investigate the evolutionary history of the LB group of spirochaetes because all loci are present in an outgroup species. The phylogenies constructed in Chapter 6 revealed a potential Eurasian origin for the species complex. Then, still early in the evolutionary history of the LB group, a subgroup of strains appear to have become isolated in North America and diverged independently of those species that remained in Europe forming several species unique to the United States. Several speciation events also occurred in quick succession in Europe. In more recent time strains of *B. burgdorferi* and *B. bissettii* have subsequently migrated back to Europe where they have since diverged from their North American forbears to form distinct European populations. Presumably this pattern reflects rare host migration events between the two continents though the precise mechanisms are not clear.

7.2.2 Migration and Spread

Host specialisation is a key process in the ecology and evolution of tick-borne zoonotic diseases. Through the use of the MLSA scheme, I have shown that host associations shape the geographic structure of the European LB species. Strains were amplified directly from ticks from England, Scotland, Wales, Latvia and Germany as well as cultured strains from France. It was observed that strains from both bird-related species investigated, *B. valaisiana* and *B. garinii*, showed evidence of spatial mixing between the countries, whilst the rodent-related *B. afzelii* showed evidence of differentiation of populations from different countries.

Further to this the *B. afzelii* phylogeny potentially indicated ancient links to its rodent hosts. The phylogeographic patterning of *B. afzelii* paralleled that of putative rodent hosts (vole and shrew species) which are believed to form clades based on ancestral refuge populations thought to have been located on the Iberian peninsular and East Baltic region during the last glacial maximum (Fumagalli *et al.*, 1996, Heckel *et al.*, 2005, Hewitt, 1999, Taberlet and Bouvet, 1994, Taberlet *et al.*, 1994). This theory was further supported by the fact that *B. afzelii* housekeeping gene sequence data suggested the population had recently undergone spatial expansion. However, estimations of date of expansion are problematic in *Borrelia* due to its varied lifestyle in nature (Hoen *et al.*, 2009). Further work calculating the age of *Borrelia* species is necessary. This could possibly be accomplished through more accurate estimations of their mutation rates and estimations of the generation time of

Borrelia during different stages of their life cycle would be invaluable to future population and evolutionary studies of the LB group spirochaetes.

I have reported for the first time *B. afzelii* in questing ticks from England and revealed the infections were highly focal. The MLSA sequence data revealed that STs found in England were also highly structured with four of the five STs from England found exclusively and often repeatedly at a single site. I have suggested that the genetically distinct localised populations represent independent and fairly recent introductions from outside the UK. Some Latvian sites also showed structuring of STs but not to the same extent as observed in England and, unlike in England, Latvian sites were all found to be harbouring more than one ST. There appeared to be genetic exchange between two of the Latvian sites while one site (Babite) remained more isolated. This structuring of *B. afzelii* strains echoed observations in rodent species. Fine scale structuring has also been observed in bank voles (*Myodes glareolus*) where natural and man made barriers were believed to reduce the rates of migration between populations (Gerlach and Musolf, 2000). Structuring in the absence of barriers also occurs in common voles (*Microtus arvalis*) were it has been suggested that social structure within populations means that dispersal from local groups is rare (Gerlach and Musolf, 2000, Schweizer *et al.*, 2007).

Sample sizes of *B. afzelii* infected ticks at sites in both England and Latvia were extremely limited and a larger scale study is recommended to confirm that the majority of *B. afzelii* strains at English sites are clonal. Furthermore, it would also be interesting to return to English sites over a longer period of time to investigate whether strains do begin to migrate over time or whether they remain isolated. It would also be interesting to screen ticks collected from areas in between neighbouring sites that have localised *B. afzelii* populations in order to identify putative barriers to migration. Furthermore, SNP analyses of *B. afzelii* strains at English sites would provide more detailed image of how diverged the strains are and may help form a better estimate of how long *B. afzelii* populations have been at the sites. Parallel studies investigating the rodent populations would be useful to directly investigate the link between structuring in rodent and *B. afzelii* populations.

7.3 LB Group Species

The large number of completed prokaryotic genome sequences has fuelled the debate about the nature of bacterial species and the need for a theory-based prokaryotic species concept became obvious. The biological species concept developed by Ernst Mayr (Mayr, 1942) states that species are groups of organisms that can interbreed and produce viable offspring. However, due to the asexual mode of reproduction and the fact that gene transfer

has been documented between highly divergent bacterial strains this concept cannot be applied to bacterial species. In an attempt to develop a more general concept that is applicable to prokaryotic species, the cohesion species concept has been proposed. This concept states that distinct bacterial (and eukaryotic) groups do exist and are maintained by selective forces based on the biological niche that the species or ecotype occupies (Achtman and Wagner, 2008, Cohan, 2002, Meglitsch, 1954). Cohan (2002) has shown that these ecotypes are remarkably congruent with sequence clusters in existence in several bacterial groups. LB spirochaetes also comprise of distinctive ecotypes defined by their spectrum of vertebrate hosts and the genetic data generated during my work also support the ecotype definition. Using the MLSA scheme I have shown that strains believed to be adapted to the same reservoir host tightly cluster at the end of long branches. In my thesis no examples of strains were found that fell outside or between these ecotypes. *B. valaisiana* and *B. garinii* are two distinct species included in my study that share the same host range and occur sympatrically (Taragel'ova *et al.*, 2008) but measures of genetic distance show that they are distinct species and not closely related and it has been proposed that these strains may have evolved allopatrically which may explain their pronounced genetic distance (Margos *et al.*, 2009).

7.4 Host Adaptation

Specialism can in many circumstances be considered a good evolutionary strategy because functional trade-offs that occur when attempting to adapt to multiple niches limit the fitness of generalists in any one habitat. Thus any specialist should out compete a generalist in any given niche (Whitlock, 1996, Woolhouse *et al.*, 2001). The majority of species specialize to the level of rodent or avian hosts but they are able to infect a broad range of species within this group so they are not true host specialists. However, some species such as *B. turdi* and *B. tanukii,* little is known about the breadth of their host range. The LB spirochaetes would be unlikely to benefit from being strict specialists because of the generalist nature of many of the vectors of the *I. persulcatus* species complex. The tick vectors are likely to dictate the selection pressures on spirochaete species through their host preference. For example, if the vector due to ecological changes or even choice altered its host range causing an increase in probability of feeding on an avian host, then this is likely to place selection pressure on the *Borrelia* species to be able to infect bird species. Therefore, prolonged changes to the ecology of a site or region may cause the host specialism of *Borrelia* species to alter. The species radiation observed in the evolutionary histories of both the LB group of spirochaetes and the *I. persulcatus* species complex may

be an example of the vector host feeding habits changing (due to ecological changes) leading to selection in the *Borrelia* species. Changes in host adaptation have occurred several times in the evolutionary history of the LB spirochaetes (usually bird or rodent) (Chapter 6).

Although genetic mechanisms of host specialization are not fully understood, a key method of host immune evasion is in the form of complement-regulator acquiring surface proteins (CRASPs), which disrupt the action of the host complement systems. It has been shown that adaptive evolution may occur through gene duplication, loss and functional diversification of CRASPs (Wywial *et al.*, 2009). Horizontal gene transfer has also been identified at plasmid located *erp* genes which are also involved in regulating host complement (Stevenson and Miller, 2003).

7.5 Host-Vector-Pathogen Coevolution

In a broad sense coevolution can be defined as the change of a biological organism triggered by the change of a related organism. Pathogen-host interactions are often used as examples of coevolution and in many cases the relationship is very close. This is exemplified by Hanta virus where each virus group shares a long standing relationship with a specific rodent species (Henttonen *et al.*, 2008). However, when the pathogen life-cycle requires vectors for transmission this adds a second dimension to the coevolutionary relationship. In the LB group of spirochaetes it is clear that the fitness of the group is dependant on both vector and host. However, as many of the major known tick vectors are generalists themselves they do not rely on specialist host groups. The hosts and vectors of the LB group of spirochaetes provide extremely different environments for survival, differing substantially in temperature and nutrient availability and the spriochaetes have adapted to both environments in order to survive. They have evolved to manipulate the vector systems for example, spriochaetes appear to increase the concentration of the TR*OSPA* receptors in the tick gut which the spirochaetes require for migration to the tick salivary glands where they are then secreted into the host (Pal *et al.*, 2004a).

My thesis has revealed potential coevolution between LB spirochaetes and the *I. persulcatus* species complex in the distant past. Trees of the LB species complex and those of the *I. persulcatus* species complex share many similarities which may represent a close evolutionary relationship in the past. However, when looking at more recent relationships between tick species and specific LB species there is no close resemblance in their evolution. As stated above and in Chapter 6, population studies have found no relationship between the population structure of *B. burgdorferi* and its main US vector, *I. scapularis*

(Humphrey *et al.*, 2010, Qiu *et al.*, 2002). My population studies revealed a close link between the population structure of small mammal hosts and *B. afzelii* suggesting a closer link with hosts and not vectors in the more recent evolutionary history of the LB species.

However, while there may be a general resemblance in the evolutionary histories of host species and LB species, the spirochaete species are not strictly specialists. *B. valaisiana* and *B. garinii* have been shown to be transmitted by mainly three different bird species; song thrushes, blackbirds and pheasants but they can also infect many others (Kurtenbach *et al.*, 1998, Taragel'ova *et al.*, 2008). Thus the correlation between host and LB species evolution highlights the dependence of the LB species on hosts species in general but does not show direct coevolution with a single host species.

The LB spirochaetes are also dependant on the *I. persulcatus* species complex for transmission between hosts. As stated in section 7.2, changes in the vector's host range will likely apply selective pressures to the *Borrelia* species to adapt to the altered host range. The radiation of tick and *Borrelia* species trees discussed in Chapter 6 may well be an example of this. It has been suggested that the radiation of tick species may have occurred in response to the increase in mammal host species after the dinosaur extinction approximately 65 million years ago (mya). It would be logical to suggest that this may have caused a knock-on effect in the LB spirochaetes. Foley and colleagues (Foley *et al.*, 2008) investigating the evolution of *Anaplasma phagocytophilum* and *A. marginale*, suggest that the speciation of these two species occurred between 78 and 43 mya and that this speciation may be due to the same diversification event in ticks and mammalian hosts.

7.6 Selection and Migration

One study has suggested that *B. garinii* and *B. valaisiana* evolved allopatrically and have since come to occupy the same geographic and host range in Europe (Margos *et al.*, 2009). However, *B. garinii* can additionally infect sea bird colonies via the tick species, *I. uriae*, and appears to have a broader geographic range, regularly being isolated in Asia (Olsen *et al.*, 1993). Gause's Law of competitive exclusion states that two species that compete for the exact same resources cannot stably coexist (Gause, 1934). It has been shown that both species not only share a bird host range but also tend to share the same host species within that range in Europe (song birds and blackbirds) (Taragel'ova *et al.*, 2008). They are also frequently found as mixed infections in the same tick suggesting they must compete, on some level, in the tick midgut (Kurtenbach *et al.*, 1998). Further research is required in this area to understand the niche differences between the species or whether it is likely one may gradually outcompete the other. I would recommend using the MLSA scheme to better

understand populations of *B. garinii* that infect sea birds. One study has attempted a preliminary examination using the 5S-23S intergenic spacer region of *B. garinii* populations in *I. uriae* (marine) and *I. ricinus* ticks (terrestrial) and found partial overlap in the populations. The MLSA scheme has been shown to be more sensitive to differences in populations and so may provide further insight into the relationship between the two cycles. It may also shed light on whether *B. garinii* in the past has moved from marine birds to terrestrial species or vice versa. Another line of investigation would be tissue tropism in the vertebrate host as it is conceivable that the two species occupy slightly different niches by infecting different tissues in the host. Further research into host species of *B. garinii* in Asia may shed light on differences in their respective niches as *B. valaisiana* appears to rarely occur on this continent (Masuzawa, 2004). In general a better understanding of the relationship between *B. garinii* and *B. valaisiana* certainly would be of clinical interest considering that *B. garinii* is often linked with the serious neuroborreliosis clinical manifestations while *B. valaisiana* has rarely been associated with disease in humans (Ornstein *et al.*, 2001, Rijpkema *et al.*, 1997).

This thesis has shown that *B. afzelii* populations in Europe are highly structured and on a finer scale it appears populations are also often highly localized. Furthermore, the preliminary data from Scottish sites suggested that there might be no migration between the north and south of Great Britain. However, while comparisons between European and Chinese *B. afzelii* populations revealed that there was substantial genetic differentiation between the two regions, it was less than expected when considering the huge geographic distance. *B. afzelii* is comparatively conserved compared to other LB species but still has one of the largest geographic distributions. No putative host species of *B. afzelii* covers the entire *B. afzelii* distribution range. In particular Chinese mouse (*Apodemus*) and vole (*Microtus* and *Myodes*) species are highly fragmented, and it is interesting that *B. afzelii* is such a closely related species when it's putative host species are fragmented (IUCN, 2010). The reasons behind this finding are not fully understood, however, it seems likely the low diversity of *B. afzelii* may be linked to a limited number of refugia during the last glacial period (Hewitt, 1999) which is likely to have forced the species through an intense bottle neck. *B. afzelii* has since rapidly recolonised Eurasia. Similarly, it was observed in Chapter 6 that the US species rarely speciated based on geographic factors and here it has also been suggested that this may be related to conditions caused by the last glacial period (Qiu *et al.*, 2002, Wiley, 1988). During the last glacial period arid and desert areas, which now act as migration barriers, were less extreme potentially allowing for migration of *Borrelia* species across the entire North American continent (Qiu *et al.*, 2002, Wiley, 1988).

However, further research is required to better understand host ranges of species across Eurasia such as *B. afzelii*. Previous work investigating host species of *B. afzelii* have revealed a highly complex picture. Bank voles (*Clethrionomys glareolus*), wood mice (*Apodemus sylvaticus*) and yellow neck mice (*Apodemus flaviollis*) have both been identified as important hosts (Hanincova *et al.*, 2003a). They do, however, vary in their susceptibility to tick vectors and ability to transmit the spirochaetes. However, the host ranges of these species do not fully cover the geographic range of *B. afzelii* (Hanincova *et al.*, 2003a, IUCN, 2010).

7.7 Future Work

My thesis has utilized a novel MLSA scheme to provide new insight into the population structure of the LB group of spirochaetes. However, there are still many areas of the LB group that are poorly understood. As stated above further work into the *B. garinii* population structure would be a worthwhile topic to further investigate. This would be especially interesting in relation to the population structure of *B. garinii* strains found in marine bird species compared to those found in terrestrial birds as well as the movement between to the two systems.

Also, further research into barriers to the movement of both bird and rodent related LB group species would also benefit our understanding of the migration and spread of this pathogen. Studies, using the MLSA scheme, are underway to investigate the effect of mountain ranges as geographic barriers, in the Austrian Alps. For *B. afzelii* further research is required to better understand which species are the most significant hosts for their transmission. Then, research is also required to better understand the migration and spread of these host species, which will in turn improve our understanding of the ability of this *Borrelia* species to migrate and spread. Work is already planned to perform blood meal analysis on ticks which uses hybridization techniques to identify what species the tick last fed on (Moran Cadenas *et al.*, 2007). Through the use of this method it will help in identifying the host species that ticks acquire LB infections and therefore, better understand the true host range of *B. afzelii*. It may also be possible to ascertain if there is a link between the make up of the tick host range and the range of *Borrelia* species present at a particular site. This information may help to predict the future migration and spread of this emerging zoonotic disease.

Further work is required to better resolve the evolution of the LB group of spirochaetes. My research suggested that the inclusion of extra genes may improve branch support and increase identity between the phylogenies. It is also recommended that genes

are more thoroughly investigated for their mode of evolution. Finally, the MLSA data should be subjected to coalescence analysis as this may assist primarily in estimating the time frame of speciation events and secondly, may be used to identify cases of incomplete lineage sorting.

7.8 Summary

During this thesis I have optimised a MLSA scheme to successfully amplify *Borrelia* DNA directly from arthropod vectors. Phylogenies made using concatenated housekeeping genes were superior to single loci approaches in several ways, including i) grouping of species, ii) detecting genetic structure within the species and iii) being able to show geographic structuring within the species. I examined the impact of host migration on the spread of a tick-borne zoonotic disease, using LB spirochaetal species in Europe. It became apparent that populations of *Borrelia* spp. associated with birds show limited structuring of populations between geographically distant regions compared to those associated with small mammals, and I argued that this can be explained by higher rates of migration in avian hosts. Further to this, the *B. afzelii* population structure mirrored that of the putative rodent host species which are believed to have been shaped by small refuge populations of the last glacial maximum. Parallels between vector group and LB spirochaete group evolution illustrated the close relationship between the vectors and pathogens. These findings highlight the absolute dependence of LB spirochaetes on both host and vector and I have suggested their complex interactions must dictate speciation events within the LB group of species. Overall, I conclude that the MLSA scheme has the potential to unravel the evolutionary and geographic origins of this bacterial tick-borne species complex as a whole.

Bibliography

ABBOTT, A. 2006. Lyme disease: uphill struggle. *Nature,* 439(7076)**,** 524-5.

ACHTMAN, M. & WAGNER, M. 2008. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol,* 6(6)**,** 431-40.

ANDERSON, J. F., MAGNARELLI, L. A. & MCANINCH, J. B. 1988. New *Borrelia burgdorferi* antigenic variant isolated from *Ixodes dammini* from upstate New York. *J Clin Microbiol,* 26(10)**,** 2209-12.

ANISIMOVA, M. & GASCUEL, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol,* 55(4)**,** 539-52.

AVISE, J. C. & ROBINSON, T. J. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Syst Biol,* 57(3)**,** 503-7.

BALASHOV, Y. S. 1972. Bloodsucking ticks (Ixodoidea) - vectors of diseases of man and animals. *Miscellaneous Publications of the Entomological Society of America,* 8163-176.

BARANTON, G., POSTIC, D., SAINT GIRONS, I., BOERLIN, P., PIFFARETTI, J. C., ASSOUS, M. & GRIMONT, P. A. 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. *Int J Syst Bacteriol,* 42(3)**,** 378-83.

BARBOUR, A. G. 1993. Linear DNA of *Borrelia* species and antigenic variation. *Trends Microbiol,* 1(6)**,** 236-9.

BARBOUR, A. G. & HAYES, S. F. 1986. Biology of *Borrelia* species. *Microbiol Rev,* 50(4)**,** 381-400.

BARBOUR, A. G., HEILAND, R. A. & HOWE, T. R. 1985. Heterogeneity of major proteins in Lyme disease *Borrelia*e: a molecular analysis of North American and European isolates. *J Infect Dis,* 152(3)**,** 478-84.

BELFAIZA, J., POSTIC, D., BELLENGER, E., BARANTON, G. & GIRONS, I. S. 1993. Genomic fingerprinting of *Borrelia burgdorferi* sensu lato by pulsed-field gel electrophoresis. *J Clin Microbiol,* 31(11)**,** 2873-7.

BERGSTROM, S., BUNDOC, V. G. & BARBOUR, A. G. 1989. Molecular analysis of linear plasmid-encoded major surface proteins, OspA and OspB, of the Lyme disease spirochaete *Borrelia burgdorferi*. *Mol Microbiol,* 3(4)**,** 479-86.

BISHOP, C. J., AANENSEN, D. M., JORDAN, G. E., KILIAN, M., HANAGE, W. P. & SPRATT, B. G. 2009. Assigning strains to bacterial species via the internet. *BMC Biol,* 73.

BISSETT, M. L. & HILL, W. 1987. Characterization of *Borrelia burgdorferi* strains isolated from *Ixodes pacificus* ticks in California. *J Clin Microbiol,* 25(12)**,** 2296-301.

BOERLIN, P., PETER, O., BRETZ, A. G., POSTIC, D., BARANTON, G. & PIFFARETTI, J. C. 1992. Population genetic analysis of *Borrelia burgdorferi* isolates by multilocus enzyme electrophoresis. *Infect Immun,* 60(4)**,** 1677-83.

BUNIKIS, J., GARPMO, U., TSAO, J., BERGLUND, J., FISH, D. & BARBOUR, A. G. 2004. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology,* 150(Pt 6)**,** 1741-55.

BUNIKIS, J., OLSEN, B., FINGERLE, V., BONNEDAHL, J., WILSKE, B. & BERGSTROM, S. 1996. Molecular polymorphism of the lyme disease agent *Borrelia garinii* in northern Europe is influenced by a novel enzootic *Borrelia* focus in the North Atlantic. *J Clin Microbiol,* 34(2)**,** 364-8.

BURGDORFER, W. 1984. Discovery of the Lyme disease spirochete and its relation to tick vectors. *Yale J Biol Med,* 57(4)**,** 515-20.

BURGDORFER, W., BARBOUR, A. G., HAYES, S. F., BENACH, J. L., GRUNWALDT, E. & DAVIS, J. P. 1982. Lyme disease-a tick-borne spirochetosis? *Science,* 216(4552)**,** 1317-9.

BUSCH, U., HIZO-TEUFEL, C., BOEHMER, R., FINGERLE, V., NITSCHKO, H., WILSKE, B. & PREAC-MURSIC, V. 1996. Three species of *Borrelia burgdorferi* sensu lato (*B. burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii*) identified from cerebrospinal fluid isolates by pulsed-field gel electrophoresis and PCR. *J Clin Microbiol,* 34(5)**,** 1072-8.

BYKOWSKI, T., WOODMAN, M. E., COOLEY, A. E., BRISSETTE, C. A., WALLICH, R., BRADE, V., KRAICZY, P. & STEVENSON, B. 2008. *Borrelia burgdorferi* complement regulator-acquiring surface proteins (BbCRASPs): Expression patterns during the mammal-tick infection cycle. *Int J Med Microbiol,* 298 Suppl 1249-56.

CANICA, M. M., NATO, F., DU MERLE, L., MAZIE, J. C., BARANTON, G. & POSTIC, D. 1993. Monoclonal antibodies for identification of *Borrelia* afzelii sp. nov. associated with late cutaneous manifestations of Lyme borreliosis. *Scand J Infect Dis,* 25(4)**,** 441-8.

CASJENS, S., PALMER, N., VAN VUGT, R., HUANG, W. M., STEVENSON, B., ROSA, P., LATHIGRA, R., SUTTON, G., PETERSON, J., DODSON, R. J., HAFT, D., HICKEY, E., GWINN, M., WHITE, O. & FRASER, C. M. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol,* 35(3)**,** 490-516.

CDC. 2008. *Summary of notifiable diseases, United States, 2007* [Online]. Atlanta: CDC. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5553a1.htm [Accessed 28/07/2010 2010].

CHU, C. Y., LIU, W., JIANG, B. G., WANG, D. M., JIANG, W. J., ZHAO, Q. M., ZHANG, P. H., WANG, Z. X., TANG, G. P., YANG, H. & CAO, W. C. 2008. Novel genospecies of *Borrelia burgdorferi* sensu lato from rodents and ticks in southwestern China. *J Clin Microbiol,* 46(9)**,** 3130-3.

COHAN, F. M. 2002. What are bacterial species? *Annu Rev Microbiol,* 56457-487.

COLLARES-PEREIRA, M., COUCEIRO, S., FRANCA, I., KURTENBACH, K., SCHAFER, S. M., VITORINO, L., GONCALVES, L., BAPTISTA, S., VIEIRA, M. L. & CUNHA, C. 2004. First isolation of *Borrelia lusitaniae* from a human patient. *J Clin Microbiol,* 42(3)**,** 1316-8.

COMSTEDT, P., ASOKLIENE, L., ELIASSON, I., OLSEN, B., WALLENSTEN, A., BUNIKIS, J. & BERGSTROM, S. 2009. Complex population structure of Lyme borreliosis group spirochete *Borrelia garinii* in subarctic Eurasia. *PLoS One,* 4(6)**,** e5841.

DANIELS, T. J. & FISH, D. 1995. Effect of deer exclusion on the abundance of immature *Ixodes scapularis* (Acari: Ixodidae) parasitizing small and medium-sized mammals. *J Med Entomol,* 32(1)**,** 5-11.

DAUTEL, H., DIPPEL, C., KAMMER, D., WERKHAUSEN, A. & KAHL, O. 2008. Winter activity of *Ixodes ricinus* in a Berlin forest. *International Journal of Medical Microbiology,* 29850-54.

DE CARVALHO, I. L., FONSECA, J. E., MARQUES, J. G., ULLMANN, A., HOJGAARD, A., ZEIDNER, N. & NUNCIO, M. S. 2008. Vasculitis-like syndrome associated with *Borrelia lusitaniae* infection. *Clin Rheumatol,* 27(12)**,** 1587-91.

DE SILVA, A. M. & FIKRIG, E. 1995. Growth and migration of *Borrelia burgdorferi* in *Ixodes* ticks during blood feeding. *Am J Trop Med Hyg,* 53(4)**,** 397-404.

DIDELOT, X. & FALUSH, D. 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics,* 175(3)**,** 1251-66.

DIUK-WASSER, M. A., GATEWOOD, A. G., CORTINAS, M. R., YAREMYCH-HAMER, S., TSAO, J., KITRON, U., HICKLING, G., BROWNSTEIN, J. S., WALKER, E., PIESMAN, J. & FISH, D. 2006. Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. *J Med Entomol,* 43(2)**,** 166-76.

DOLAN, M. C., PIESMAN, J., MBOW, M. L., MAUPIN, G. O., PETER, O., BROSSARD, M. & GOLDE, W. T. 1998. Vector competence of Ixodes scapularis and *Ixodes ricinus* (Acari: Ixodidae) for three genospecies of *Borrelia burgdorferi*. *J Med Entomol,* 35(4)**,** 465-70.

DUBSKA, L., LITERAK, I., KOCIANOVA, E., TARAGELOVA, V. & SYCHRA, O. 2009. Differential role of passerine birds in distribution of *Borrelia* spirochetes, based on data from ticks collected from birds during the postbreeding migration period in Central Europe. *Appl Environ Microbiol,* 75(3)**,** 596-602.

DUNEAU, D., BOULINIER, T., GOMEZ-DIAZ, E., PETERSEN, A., TVERAA, T., BARRETT, R. T. & MCCOY, K. D. 2008. Prevalence and diversity of Lyme borreliosis bacteria in marine birds. *Infect Genet Evol,* 8(3)**,** 352-9.

DYKHUIZEN, D. E., POLIN, D. S., DUNN, J. J., WILSKE, B., PREAC-MURSIC, V., DATTWYLER, R. J. & LUFT, B. J. 1993. *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci U S A,* 90(21)**,** 10163-7.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32(5)**,** 1792-7.

EISEN, L. & LANE, R. S. 2002a. Vectors of *Borrelia burgdorferi* sensu lato. *In:* GRAY, J. S., KAHL, O., LANE, R. S. & STANEK, G. (eds.) *Lyme Borreliosis: Biology, Epidemiology and Control.* Wallingford, UK: CABI Publishing.

EISEN, L. & LANE, R. S. 2002b. Vectors of *Borrelia burgdorferi* sensu lato. *In:* LANE, R. S. & STANEK, G. (eds.) *Lyme Borreliosis: Biology, Epidemiology and Control.* Wallingford, UK: CABI Publishing.

ESTRADA-PENA, A., OSACAR, J. J., PICHON, B. & GRAY, J. S. 2005. Hosts and pathogen detection for immature stages of *Ixodes ricinus* (Acari: *Ixodidae*) in North-Central Spain. *Exp Appl Acarol,* 37(3-4)**,** 257-68.

ETTI, S., HAILS, R., SCHAFER, S. M., DE MICHELIS, S., SEWELL, H. S., BORMANE, A., DONAGHY, M. & KURTENBACH, K. 2003. Habitat-specific diversity of *Borrelia burgdorferi* sensu lato in Europe, exemplified by data from Latvia. *Appl Environ Microbiol,* 69(5)**,** 3008-10.

EXCOFFIER, L., LAVAL, G. & SCHNEIDER, S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online,* 147-50.

FALCO, R. C. & FISH, D. 1991. Horizontal movement of adult *Ixodes dammini* (Acari: Ixodidae) attracted to CO2-baited traps. *J Med Entomol,* 28(5)**,** 726-9.

FEIL, E. J., ENRIGHT, M. C. & SPRATT, B. G. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between Neisseria meningitidis and Streptococcus pneumoniae. *Res Microbiol,* 151(6)**,** 465-9.

FEIL, E. J., MAIDEN, M. C., ACHTMAN, M. & SPRATT, B. G. 1999. The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis. *Mol Biol Evol,* 16(11)**,** 1496-502.

FELSENSTEIN, J. 2005. *PHYLIP (Phylogeny Inference Package)* [Online]. Seattle: Felsenstein, J. Available: http://evolution.genetics.washington.edu/phylip/faq.html#citation [Accessed 12 August 2010 2010].

FILIPPOVA, N. A. 1973. [Species of the group *Ixodes persulcatus* (Parasitiformes, Ixodidae). VII. Paleogenesis of the southern branch of the group *Ixodes ricinus* (L.)]. *Parazitologiia,* 7(1)**,** 3-13.

FILIPPOVA, N. A. 1991. A hypothesis for the palaeogenesis of the distribution of the main vectors. *In:* DUSBABEK, F. & BURKVA, V. (eds.) *Modern Acarology.* Prague: Prague and SPB Academic Publishing.

FINGERLE, V., LAUX, H., MUNDERLOH, U. G., SCHULTE-SPECHTEL, U. & WILSKE, B. 2000. Differential expression of outer surface proteins A and C by individual *Borrelia burgdorferi* in different genospecies. *Med Microbiol Immunol,* 189(2)**,** 59-66.

FINGERLE, V., MICHEL, H., HETTCHE, G., HIZO-TEUFEL, C. & WILSKE, B. 2004. *Borrelia burgdorferi* s.l. OspA-types are widespread in Bavaria but show distinct local patterns. *Int J Med Microbiol,* 293 Suppl 37165-6.

FINGERLE, V., SCHULTE-SPECHTEL, U. C., RUZIC-SABLJIC, E., LEONHARD, S., HOFMANN, H., WEBER, K., PFISTER, K., STRLE, F. & WILSKE, B. 2008. Epidemiological aspects and molecular characterization of *Borrelia burgdorferi* s.l. from southern Germany with special respect to the new species *Borrelia spielmanii* sp. nov. *Int J Med Microbiol,* 298(3-4)**,** 279-90.

FITZPATRICK, D. A. & MCINERNEY, J. O. 2005. Evidence of positive Darwinian selection in Omp85, a highly conserved bacterial outer membrane protein essential for cell viability. *J Mol Evol,* 60(2)**,** 268-73.

FOLDVARI, G., FARKAS, R. & LAKOS, A. 2005. *Borrelia spielmanii* erythema migrans, Hungary. *Emerg Infect Dis,* 11(11)**,** 1794-5.

FOLEY, J., NIETO, N. C., FOLEY, P. & TEGLAS, M. B. 2008. Co-phylogenetic analysis of *Anaplasma phagocytophilum* and its vectors, *Ixodes* spp. ticks. *Exp Appl Acarol,* 45(3-4)**,** 155-70.

FORETZ, M., POSTIC, D. & BARANTON, G. 1997. Phylogenetic analysis of *Borrelia burgdorferi* sensu stricto by arbitrarily primed PCR and pulsed-field gel electrophoresis. *Int J Syst Bacteriol,* 47(1)**,** 11-8.

FRANCISCO, A. P., BUGALHO, M., RAMIREZ, M. & CARRICO, J. A. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics,* 10152.

FRASER, C. M., CASJENS, S., HUANG, W. M., SUTTON, G. G., CLAYTON, R., LATHIGRA, R., WHITE, O., KETCHUM, K. A., DODSON, R., HICKEY, E. K., GWINN, M., DOUGHERTY, B., TOMB, J. F., FLEISCHMANN, R. D., RICHARDSON, D., PETERSON, J., KERLAVAGE, A. R., QUACKENBUSH, J., SALZBERG, S., HANSON, M., VAN VUGT, R., PALMER, N., ADAMS, M. D., GOCAYNE, J., WEIDMAN, J., UTTERBACK, T., WATTHEY, L., MCDONALD, L., ARTIACH, P., BOWMAN, C., GARLAND, S., FUJI, C., COTTON, M. D., HORST, K., ROBERTS, K., HATCH, B., SMITH, H. O. & VENTER, J. C. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature,* 390(6660)**,** 580-6.

FREELAND, J. R. 2005. *Molecular Ecology,* Chichester, Wiley Blackwell.

FUKUNAGA, M., HAMASE, A., OKADA, K., INOUE, H., TSURUTA, Y., MIYAMOTO, K. & NAKAO, M. 1996a. Characterization of spirochetes isolated from ticks *(Ixodes tanuki, Ixodes turdus*, and *Ixodes columnae*) and comparison of the sequences with those of *Borrelia burgdorferi* sensu lato strains. *Appl Environ Microbiol,* 62(7)**,** 2338-44.

FUKUNAGA, M., HAMASE, A., OKADA, K. & NAKAO, M. 1996b. *Borrelia* tanukii sp. nov. and *Borrelia* turdae sp. nov. found from *Ixodid* ticks in Japan: rapid species identification by 16S rRNA gene-targeted PCR analysis. *Microbiol Immunol,* 40(11)**,** 877-81.

FUKUNAGA, M., YABUKI, M., HAMASE, A., OLIVER, J. H., JR. & NAKAO, M. 2000. Molecular phylogenetic analysis of Ixodid ticks based on the ribosomal DNA spacer, internal transcribed spacer 2, sequences. *J Parasitol,* 86(1)**,** 38-43.

FUMAGALLI, L., HAUSSER, J., TABERLET, P., GIELLY, L. & STEWART, D. T. 1996. Phylogenetic structures of the Holarctic *Sorex araneus* group and its relationships with *S. samniticus*, as inferred from nitDNA sequences. *Hereditas,* 125191-199.

GAUSE, G. F. 1934. *The struggle for existence,* Baltimore, The Williams and Wilkins Company.

GERLACH, G. & MUSOLF, K. 2000. Fragmentation of Landscape as a cause for genetic subdivision in bank voles. *Conservation Biology,* 14(4)**,** 1066-1074.

GERN, L. 2008. *Borrelia burgdorferi* sensu lato, the agent of lyme borreliosis: life in the wilds. *Parasite,* 15(3)**,** 244-7.

GERN, L. & RAIS, O. 1996. Efficient transmission of *Borrelia burgdorferi* between cofeeding *Ixodes ricinus* ticks (Acari: Ixodidae). *J Med Entomol,* 33(1)**,** 189-92.

GEVERS, D., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMME, P., THOMPSON, F. L. & SWINGS, J. 2005. Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol,* 3(9)**,** 733-9.

GINSBERG, H. S., BUCKLEY, P. A., BALMFORTH, M. G., ZHIOUA, E., MITRA, S. & BUCKLEY, F. G. 2005. Reservoir competence of native North American birds for the lyme disease spirochete, *Borrelia burgdorferi*. *J Med Entomol,* 42(3)**,** 445-9.

GIRARD, Y. A., TRAVINSKY, B., SCHOTTHOEFER, A., FEDOROVA, N., EISEN, R. J., EISEN, L., BARBOUR, A. G. & LANE, R. S. 2009. Population structure of the Lyme borreliosis spirochete *Borrelia burgdorferi* in the western black-legged tick (*Ixodes pacificus*) in Northern California. *Appl Environ Microbiol,* 75(22)**,** 7243-52.

GLOCKNER, G., LEHMANN, R., ROMUALDI, A., PRADELLA, S., SCHULTE-SPECHTEL, U., SCHILHABEL, M., WILSKE, B., SUHNEL, J. & PLATZER, M. 2004. Comparative analysis of the *Borrelia garinii* genome. *Nucleic Acids Res,* 32(20)**,** 6038-46.

GLOCKNER, G., SCHULTE-SPECHTEL, U., SCHILHABEL, M., FELDER, M., SUHNEL, J., WILSKE, B. & PLATZER, M. 2006. Comparative genome analysis: selection pressure on the *Borrelia* vls cassettes is essential for infectivity. *BMC Genomics,* 7211.

GOLDSTEIN, S. F., BUTTLE, K. F. & CHARON, N. W. 1996. Structural analysis of the *Leptospiraceae* and *Borrelia burgdorferi* by high-voltage electron microscopy. *J Bacteriol,* 178(22)**,** 6539-45.

GOLDSTEIN, S. F., CHARON, N. W. & KREILING, J. A. 1994. *Borrelia burgdorferi* swims with a planar waveform similar to that of eukaryotic flagella. *Proc Natl Acad Sci U S A,* 91(8)**,** 3433-7.

GRANWEHR, B. P., LILLIBRIDGE, K. M., HIGGS, S., MASON, P. W., ARONSON, J. F., CAMPBELL, G. A. & BARRETT, A. D. 2004. West Nile virus: where are we now? *Lancet Infect Dis,* 4(9)**,** 547-56.

GRAY, J. S., KAHL, O., JANETZKI, C. & STEIN, J. 1992. Studies on the ecology of Lyme disease in a deer forest in County Galway, Ireland. *J Med Entomol,* 29(6)**,** 915-20.

GRAY, J. S., KIRSTEIN, F., ROBERTSON, J. N., STEIN, J. & KAHL, O. 1999. *Borrelia burgdorferi* sensu lato in *Ixodes ricinus* ticks and rodents in a recreational park in south-western Ireland. *Exp Appl Acarol,* 23(9)**,** 717-29.

GREGO, E., BERTOLOTTI, L., PELETTO, S., AMORE, G., TOMASSONE, L. & MANNELLI, A. 2007. *Borrelia* lusitaniae *OspA* gene heterogeneity in Mediterranean basin area. *J Mol Evol,* 65(5)**,** 512-8.

GUINDON, S. & GASCUEL, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol,* 52(5)**,** 696-704.
GUY, E. C. & STANEK, G. 1991. Detection of *Borrelia burgdorferi* in patients with Lyme disease by the polymerase chain reaction. *J Clin Pathol,* 44(7)**,** 610-1.
HALL, B. G. 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Mol Biol Evol,* 22(3)**,** 792-802.
HALL, B. G. 2008. *Phylogenetic trees made easy: A how-to manual,* Third Edition. Sunderland, Massachusetts, USA, Sinauer Associates.
HANAGE, W. P., FRASER, C. & SPRATT, B. G. 2006. Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci,* 361(1475)**,** 1917-27.
HANINCOVA, K., KURTENBACH, K., DIUK-WASSER, M., BREI, B. & FISH, D. 2006. Epidemic spread of Lyme borreliosis, northeastern United States. *Emerg Infect Dis,* 12(4)**,** 604-11.
HANINCOVA, K., LIVERIS, D., SANDIGURSKY, S., WORMSER, G. P. & SCHWARTZ, I. 2008. *Borrelia burgdorferi* sensu stricto is clonal in patients with early Lyme borreliosis. *Appl Environ Microbiol,* 74(16)**,** 5008-14.
HANINCOVA, K., SCHAFER, S. M., ETTI, S., SEWELL, H. S., TARAGELOVA, V., ZIAK, D., LABUDA, M. & KURTENBACH, K. 2003a. Association of *Borrelia afzelii* with rodents in Europe. *Parasitology,* 126(Pt 1)**,** 11-20.
HANINCOVA, K., TARAGELOVA, V., KOCI, J., SCHAFER, S. M., HAILS, R., ULLMANN, A. J., PIESMAN, J., LABUDA, M. & KURTENBACH, K. 2003b. Association of *Borrelia garinii* and *B. valaisiana* with songbirds in Slovakia. *Appl Environ Microbiol,* 69(5)**,** 2825-30.
HECKEL, G., BURRI, R., FINK, S., DESMET, J. F. & EXCOFFIER, L. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution,* 59(10)**,** 2231-42.
HENTTONEN, H., BUCHY, P., SUPUTTAMONGKOL, Y., JITTAPALAPONG, S., HERBRETEAU, V., LAAKKONEN, J., CHAVAL, Y., GALAN, M., DOBIGNY, G., CHARBONNEL, N., MICHAUX, J., COSSON, J. F., MORAND, S. & HUGOT, J. P. 2008. Recent discoveries of new hantaviruses widen their range and question their origins. *Ann N Y Acad Sci,* 114984-9.
HEWITT, G. M. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society,* 6887-112.
HILLYARD, P. D. 1996. *Ticks of North-West Europe,* London, Field Studies Council.
HOEN, A. G., MARGOS, G., BENT, S. J., DIUK-WASSER, M. A., BARBOUR, A., KURTENBACH, K. & FISH, D. 2009. Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. *Proc Natl Acad Sci U S A,* 106(35)**,** 15013-8.
HORDIJK, W. & GASCUEL, O. 2005. Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics,* 21(24)**,** 4338-47.
HPA. 2010a. *Diagnosis and Treatment of Lyme borreliosis* [Online]. London: Health Protection Agency. Available: http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/LymeDisease/Guidelines/lymDiagnosisofLymeborreliosis/ [Accessed 11 August 2010 2010].
HPA. 2010b. *Epidemiology of Lyme borreliosis in the UK* [Online]. Southampton: Health Protection Agency. Available: http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1195733752250?p=1160495617097 [Accessed 4 Feb 2010 2010].
HUEGLI, D., HU, C. M., HUMAIR, P. F., WILSKE, B. & GERN, L. 2002. *Apodemus* species mice are reservoir hosts of *Borrelia garinii* OspA serotype 4 in Switzerland. *J Clin Microbiol,* 40(12)**,** 4735-7.

HUELSENBECK, J. P., BULL, J. J. & CUNNINGHAM, C. W. 1996. Combining data in phylogenetic analysis. *Tree,* 11(4)**,** 152-158.
HUMPHREY, P. T., CAPORALE, D. A. & BRISSON, D. 2010. Uncoordinated Phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes Scapularis. Evolution*.
HUSON, D. H. & BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol,* 23(2)**,** 254-67.
HUSON, D. H., RICHTER, D. C., RAUSCH, C., DEZULIAN, T., FRANZ, M. & RUPP, R. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics,* 8460.
IUCN. 2010. *IUCN Red List of Threatened Species* [Online]. Cambridge: International Union for Conservation of Nature and Natural Resources. Available: www.iucnredlist.org [Accessed 08 August 2010 2010].
JAURIS-HEIPKE, S., LIEGL, G., PREAC-MURSIC, V., ROSSLER, D., SCHWAB, E., SOUTSCHEK, E., WILL, G. & WILSKE, B. 1995. Molecular analysis of genes encoding outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato: relationship to *ospA* genotype and evidence of lateral gene exchange of *ospC*. *J Clin Microbiol,* 33(7)**,** 1860-6.
JOHNSEN, R. C., SCHMID, G. P., HYDE, F. W., STEIGERWALT, A. G. & BRENNER, D. J. 1984. *Borrelia burgdorferi* sp. nov.: etiological agent of Lyme disease. *International Journal of Systematic Bacteriology,* 34496-497.
KAWABATA, H., MASUZAWA, T. & YANAGIHARA, Y. 1993. Genomic analysis of *Borrelia japonica* sp. nov. isolated from Ixodes ovatus in Japan. *Microbiol Immunol,* 37(11)**,** 843-8.
KLOMPEN, J. S., BLACK, W. C. T., KEIRANS, J. E. & OLIVER, J. H., JR. 1996. Evolution of ticks. *Annu Rev Entomol,* 41141-61.
KNAP, N., DURMISI, E., SAKSIDA, A., KORVA, M., PETROVEC, M. & AVSIC-ZUPANC, T. 2009. Influence of climatic factors on dynamics of questing *Ixodes ricinus* ticks in Slovenia. *Vet Parasitol,* 164(2-4)**,** 275-81.
KRAICZY, P., SKERKA, C., BRADE, V. & ZIPFEL, P. F. 2001. Further characterization of complement regulator-acquiring surface proteins of *Borrelia burgdorferi*. *Infect Immun,* 69(12)**,** 7800-9.
KUMAR, S., TAMURA, K. & NEI, M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform,* 5(2)**,** 150-63.
KURTENBACH, K., DE MICHELIS, S., ETTI, S., SCHAFER, S. M., SEWELL, H. S., BRADE, V. & KRAICZY, P. 2002a. Host association of *Borrelia burgdorferi* sensu lato--the key role of host complement. *Trends Microbiol,* 10(2)**,** 74-9.
KURTENBACH, K., DE MICHELIS, S., SEWELL, H. S., ETTI, S., SCHAFER, S. M., HAILS, R., COLLARES-PEREIRA, M., SANTOS-REIS, M., HANINCOVA, K., LABUDA, M., BORMANE, A. & DONAGHY, M. 2001. Distinct combinations of *Borrelia burgdorferi* sensu lato genospecies found in individual questing ticks from Europe. *Appl Environ Microbiol,* 67(10)**,** 4926-9.
KURTENBACH, K., DE MICHELIS, S., SEWELL, H. S., ETTI, S., SCHAFER, S. M., HOLMES, E., HAILS, R., COLLARES-PEREIRA, M., SANTOS-REIS, M., HANINCOVA, K., LABUDA, M., BORMANE, A. & DONAGHY, M. 2002b. The key roles of selection and migration in the ecology of Lyme borreliosis. *Int J Med Microbiol,* 291 Suppl 33152-4.
KURTENBACH, K., GATEWOOD HOEN, A. G., BENT, S. J., VOLLMER, S. A., OGDEN, N. H. & MARGOS, G. 2010. Population biology of Lyme Borreliosis spirochetes. *In:* ROBINSON, D. A., FALUSH, D. & FEIL, E. J. (eds.) *Bacterial population genetics in infectious disease.* Hoboken, New Jersey: Wiley-Blackwell.

KURTENBACH, K., HANINCOVA, K., TSAO, J. I., MARGOS, G., FISH, D. & OGDEN, N. H. 2006. Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nat Rev Microbiol,* 4(9)**,** 660-9.

KURTENBACH, K., KAMPEN, H., DIZIJ, A., ARNDT, S., SEITZ, H. M., SCHAIBLE, U. E. & SIMON, M. M. 1995. Infestation of rodents with larval *Ixodes ricinus* (Acari: Ixodidae) is an important factor in the transmission cycle of *Borrelia burgdorferi* s.l. in German woodlands. *J Med Entomol,* 32(6)**,** 807-17.

KURTENBACH, K., PEACEY, M., RIJPKEMA, S. G., HOODLESS, A. N., NUTTALL, P. A. & RANDOLPH, S. E. 1998. Differential transmission of the genospecies of *Borrelia burgdorferi* sensu lato by game birds and small rodents in England. *Appl Environ Microbiol,* 64(4)**,** 1169-74.

KURTENBACH, K., SCHAFER, S. M., DE MICHELIS, S., ETTI, S. & SEWELL, H. S. 2002c. *Borrelia burgdorferi* sensu lato in the vertebrate host. *In:* GRAY, J. S., KAHL, O., LANE, R. S. & STANEK, G. (eds.) *Lyme Borreliosis: biology, epidemiology and control.* New York: CABI Publishing.

LANE, R. S., BROWN, R. N., PIESMAN, J. & PEAVEY, C. A. 1994. Vector competence of Ixodes pacificus and Dermacentor occidentalis (Acari: Ixodidae) for various isolates of Lyme disease spirochetes. *J Med Entomol,* 31(3)**,** 417-24.

LARSSON, C., COMSTEDT, P., OLSEN, B. & BERGSTROM, S. 2007. First record of Lyme disease *Borrelia* in the Arctic. *Vector Borne Zoonotic Dis,* 7(3)**,** 453-6.

LASTAVICA, C. C., WILSON, M. L., BERARDI, V. P., SPIELMAN, A. & DEBLINGER, R. D. 1989. Rapid emergence of a focal epidemic of Lyme disease in coastal Massachusetts. *N Engl J Med,* 320(3)**,** 133-7.

LE FLECHE, A., POSTIC, D., GIRARDET, K., PETER, O. & BARANTON, G. 1997. Characterization of *Borrelia lusitaniae* sp. nov. by 16S ribosomal DNA sequence analysis. *Int J Syst Bacteriol,* 47(4)**,** 921-5.

LEES, A. D. & MILNE, A. 1951. The seasonal and diurnal activities of individual sheep ticks (*Ixodes ricinus* L.). *Parasitology,* 26489-500.

LENCAKOVA, D., HIZO-TEUFEL, C., PETKO, B., SCHULTE-SPECHTEL, U., STANKO, M., WILSKE, B. & FINGERLE, V. 2006. Prevalence of *Borrelia burgdorferi* s.l. OspA types in *Ixodes ricinus* ticks from selected localities in Slovakia and Poland. *Int J Med Microbiol,* 296 Suppl 40108-18.

LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25(11)**,** 1451-2.

LIEBISCH, G., SOHNS, B. & BAUTSCH, W. 1998. Detection and typing of *Borrelia burgdorferi* sensu lato in *Ixodes ricinus* ticks attached to human skin by PCR. *J Clin Microbiol,* 36(11)**,** 3355-8.

LIN, T., OLIVER, J. H., JR. & GAO, L. 2002. Genetic diversity of the outer surface protein C gene of southern *Borrelia* isolates and its possible epidemiological, clinical, and pathogenetic implications. *J Clin Microbiol,* 40(7)**,** 2572-83.

LINDGREN, E. & JAENSON, T. G. 2006. Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and measures. Copenhagen: World Health Organization.

LING, C. L., JOSS, A. W., DAVIDSON, M. M. & HO-YEN, D. O. 2000. Identification of different *Borrelia burgdorferi* genomic groups from Scottish ticks. *Mol Pathol,* 53(2)**,** 94-8.

LIVERIS, D., WORMSER, G. P., NOWAKOWSKI, J., NADELMAN, R., BITTKER, S., COOPER, D., VARDE, S., MOY, F. H., FORSETER, G., PAVIA, C. S. & SCHWARTZ, I. 1996. Molecular typing of *Borrelia burgdorferi* from Lyme disease patients by PCR-restriction fragment length polymorphism analysis. *J Clin Microbiol,* 34(5)**,** 1306-9.

MADDISON, W. P. 1997. Gene Trees in Species Trees. *Systematic Biology,* 46(3)**,** 523-536.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A,* 95(6)**,** 3140-5.

MANEL, S., SCHWARTZ, M. K., LUIKART, G. & RABERLET, P. 2003. Landscape genetics: combining landscape ecology and populations genetics. *Trends Ecology and Evolution,* 18189-197.

MARCONI, R. T. & GARON, C. F. 1992. Development of polymerase chain reaction primer sets for diagnosis of Lyme disease and for species-specific identification of Lyme disease isolates by 16S rRNA signature nucleotide analysis. *J Clin Microbiol,* 30(11)**,** 2830-4.

MARCONI, R. T., HOHENBERGER, S., JAURIS-HEIPKE, S., SCHULTE-SPECHTEL, U., LAVOIE, C. P., ROSSLER, D. & WILSKE, B. 1999. Genetic analysis of *Borrelia garinii* OspA serotype 4 strains associated with neuroborreliosis: evidence for extensive genetic homogeneity. *J Clin Microbiol,* 37(12)**,** 3965-70.

MARCONI, R. T., LIVERIS, D. & SCHWARTZ, I. 1995. Identification of novel insertion elements, restriction fragment length polymorphism patterns, and discontinuous 23S rRNA in Lyme disease spirochetes: phylogenetic analyses of rRNA genes and their intergenic spacers in *Borrelia japonica* sp. nov. and genomic group 21038 (*Borrelia andersonii* sp. nov.) isolates. *J Clin Microbiol,* 33(9)**,** 2427-34.

MARGOS, G., GATEWOOD, A. G., AANENSEN, D. M., HANINCOVA, K., TEREKHOVA, D., VOLLMER, S. A., CORNET, M., PIESMAN, J., DONAGHY, M., BORMANE, A., HURN, M. A., FEIL, E. J., FISH, D., CASJENS, S., WORMSER, G. P., SCHWARTZ, I. & KURTENBACH, K. 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A,* 105(25)**,** 8730-5.

MARGOS, G., VOLLMER, S. A., CORNET, M., GARNIER, M., FINGERLE, V., WILSKE, B., BORMANE, A., VITORINO, L., COLLARES-PEREIRA, M., DRANCOURT, M. & KURTENBACH, K. 2009. A new *Borrelia* species defined by multilocus sequence analysis of housekeeping genes. *Appl Environ Microbiol,* 75(16)**,** 5410-6.

MARTI RAS, N. M., POSTIC, D., FORETZ, M. & BARANTON, G. 1997. *Borreila burgdorferi* sensu stricto, a bacterial species "made in the U.S.A."? *International Journal of Systematic Bacteriology,* 47(4)**,** 1112-1117.

MASUZAWA, T. 2004. Terrestrial distribution of the Lyme borreliosis agent *Borrelia burgdorferi* sensu lato in East Asia. *Jpn J Infect Dis,* 57(6)**,** 229-35.

MASUZAWA, T., HASHIMOTO, N., KUDEKEN, M., KADOSAKA, T., NAKAMURA, M., KAWABATA, H., KOIZUMI, N. & IMAI, Y. 2004. New genomospecies related to *Borrelia valaisiana*, isolated from mammals in Okinawa archipelago, Japan. *J Med Microbiol,* 53(Pt 5)**,** 421-6.

MASUZAWA, T., TAKADA, N., KUDEKEN, M., FUKUI, T., YANO, Y., ISHIGURO, F., KAWAMURA, Y., IMAI, Y. & EZAKI, T. 2001. *Borrelia* sinica sp. nov., a lyme disease-related *Borrelia* species isolated in China. *Int J Syst Evol Microbiol,* 51(Pt 5)**,** 1817-24.

MATHIESEN, D. A., OLIVER, J. H., JR., KOLBERT, C. P., TULLSON, E. D., JOHNSON, B. J., CAMPBELL, G. L., MITCHELL, P. D., REED, K. D., TELFORD, S. R., 3RD, ANDERSON, J. F., LANE, R. S. & PERSING, D. H. 1997. Genetic heterogeneity of *Borrelia burgdorferi* in the United States. *J Infect Dis,* 175(1)**,** 98-107.

MAYR, E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist,* Cambridge, MA, Harvard University Press.

MEDIANNIKOV, O. Y., IVANOV, L., ZDANOVSKAYA, N., VOROBYOVA, R., SIDELNIKOV, Y., FOURNIER, P. E., TARASEVICH, I. & RAOULT, D. 2005. Diversity of *Borrelia burgdorferi* sensu lato in Russian Far East. *Microbiol Immunol,* 49(3)**,** 191-7.

MEDLOCK, J. M., PIETZSCH, M. E., RICE, N. V., JONES, L., KERROD, E., AVENELL, D., LOS, S., RATCLIFFE, N., LEACH, S. & BUTT, T. 2008. Investigation of ecological and environmental determinants for the presence of questing *Ixodes ricinus* (Acari: Ixodidae) on Gower, South Wales. *J Med Entomol,* 45(2)**,** 314-25.

MEGLITSCH, P. A. 1954. On the nature of species. *Syst Zool,* 3491-503.

METOFFICE. 2010. *Met Office: UK actual and anomaly maps* [Online]. Exeter: The Met Office. Available: http://www.metoffice.gov.uk/climate/uk/anomalygraphs/index.html# [Accessed 9 Mar 2010 2010].

MILNE, A. 1944. The ecology of the sheep tick, *Ixodes ricinus* L.: distribution of the tick in relation to geology, soil and vegetation in northern England. *Parasitology,* 35186-196.

MORAN CADENAS, F., RAIS, O., HUMAIR, P. F., DOUET, V., MORET, J. & GERN, L. 2007. Identification of host bloodmeal source and *Borrelia burgdorferi* sensu lato in field-collected *Ixodes ricinus* ticks in Chaumont (Switzerland). *J Med Entomol,* 44(6)**,** 1109-17.

NADELMAN, R. B., NOWAKOWSKI, J., FORSETER, G., GOLDBERG, N. S., BITTKER, S., COOPER, D., AGUERO-ROSENFELD, M. & WORMSER, G. P. 1996. The clinical spectrum of early Lyme borreliosis in patients with culture-confirmed erythema migrans. *Am J Med,* 100(5)**,** 502-8.

NAKAO, M. & SATO, Y. 1996. Refeeding activity of immature ticks of *Ixodes persulcatus* and transmission of Lyme disease spirochete by partially fed larvae. *J Parasitol,* 82(4)**,** 669-72.

NORRIS, S. J., HOWELL, J. K., GARZA, S. A., FERDOWS, M. S. & BARBOUR, A. G. 1995. High- and low-infectivity phenotypes of clonal populations of in vitro-cultured *Borrelia burgdorferi*. *Infect Immun,* 63(6)**,** 2206-12.

OGDEN, N. H., LINDSAY, L. R., HANINCOVA, K., BARKER, I. K., BIGRAS-POULIN, M., CHARRON, D. F., HEAGY, A., FRANCIS, C. M., O'CALLAGHAN, C. J., SCHWARTZ, I. & THOMPSON, R. A. 2008. Role of migratory birds in introduction and range expansion of *Ixodes scapularis* ticks and of *Borrelia burgdorferi* and Anaplasma phagocytophilum in Canada. *Appl Environ Microbiol,* 74(6)**,** 1780-90.

OGDEN, T. H. & ROSENBERG, M. S. 2006. Multiple sequence alignment accuracy and phylogenetic inference. *Syst Biol,* 55(2)**,** 314-28.

OJAIMI, C., DAVIDSON, B. E., SAINT GIRONS, I. & OLD, I. G. 1994. Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, B. *garinii* and B. afzelii. *Microbiology,* 140 (Pt 11)2931-40.

OLSEN, B., JAENSON, T. G. & BERGSTROM, S. 1995. Prevalence of *Borrelia burgdorferi* sensu lato-infected ticks on migrating birds. *Appl Environ Microbiol,* 61(8)**,** 3082-7.

OLSEN, B., JAENSON, T. G., NOPPA, L., BUNIKIS, J. & BERGSTROM, S. 1993. A Lyme borreliosis cycle in seabirds and *Ixodes uriae* ticks. *Nature,* 362(6418)**,** 340-2.

ORNSTEIN, K., BERGLUND, J., NILSSON, I., NORRBY, R. & BERGSTROM, S. 2001. Characterization of Lyme borreliosis isolates from patients with erythema migrans and neuroborreliosis in southern Sweden. *J Clin Microbiol,* 39(4)**,** 1294-8.

OSTFELD, R. S., CANHAM, C. D., OGGENFUSS, K., WINCHCOMBE, R. J. & KEESING, F. 2006. Climate, deer, rodents, and acorns as determinants of variation in lyme-disease risk. *PLoS Biol,* 4(6)**,** e145.

OSTFELD, R. S., SCHAUBER, E. M., CANHAM, C. D., KEESING, F., JONES, C. G. & WOLFF, J. O. 2001. Effects of acorn production and mouse abundance on abundance and *Borrelia burgdorferi* infection prevalence of nymphal Ixodes scapularis ticks. *Vector Borne Zoonotic Dis,* 1(1)**,** 55-63.

PAGE, R. D. M. & HOLMES, E. C. 1998. *Molecular evolution a phylogenetic approach,* 11. Oxford, Blackwell Publishing Ltd.

PAL, U., LI, X., WANG, T., MONTGOMERY, R. R., RAMAMOORTHI, N., DESILVA, A. M., BAO, F., YANG, X., PYPAERT, M., PRADHAN, D., KANTOR, F. S., TELFORD, S., ANDERSON, J. F. & FIKRIG, E. 2004a. TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. *Cell,* 119(4)**,** 457-68.

PAL, U., YANG, X., CHEN, M., BOCKENSTEDT, L. K., ANDERSON, J. F., FLAVELL, R. A., NORGARD, M. V. & FIKRIG, E. 2004b. OspC facilitates *Borrelia burgdorferi* invasion of Ixodes scapularis salivary glands. *J Clin Invest,* 113(2)**,** 220-30.

PECCHIOLI, E., HAUFFE, H. C., TAGLIAPIETRA, V., BANDI, C., GENCHI, C. & RIZZOLI, A. 2007. Genospecies of *Borrelia burgdorferi* sensu lato in *Ixodes ricinus* ticks from the Autonomous Province of Trento, Italy. *Int J Med Microbiol,* 297(1)**,** 53-9.

PIESMAN, J. & GERN, L. 2004. Lyme borreliosis in Europe and North America. *Parasitology,* 129 SupplS191-220.

PIETZSCH, M. E., MEDLOCK, J. M., JONES, L., AVENELL, D., ABBOTT, J., HARDING, P. & LEACH, S. 2005. Distribution of *Ixodes ricinus* in the British Isles: investigation of historical records. *Med Vet Entomol,* 19(3)**,** 306-14.

POSTIC, D., ASSOUS, M. V., GRIMONT, P. A. & BARANTON, G. 1994. Diversity of *Borrelia burgdorferi* sensu lato evidenced by restriction fragment length polymorphism of rrf (5S)-rrl (23S) intergenic spacer amplicons. *Int J Syst Bacteriol,* 44(4)**,** 743-52.

POSTIC, D., EDLINGER, C., RICHAUD, C., GRIMONT, F., DUFRESNE, Y., PEROLAT, P., BARANTON, G. & GRIMONT, P. A. 1990. Two genomic species in *Borrelia burgdorferi*. *Res Microbiol,* 141(4)**,** 465-75.

POSTIC, D., GARNIER, M. & BARANTON, G. 2007. Multilocus sequence analysis of atypical *Borrelia burgdorferi* sensu lato isolates--description of *Borrelia californiensis* sp. nov., and genomospecies 1 and 2. *Int J Med Microbiol,* 297(4)**,** 263-71.

POSTIC, D., RAS, N. M., LANE, R. S., HENDSON, M. & BARANTON, G. 1998. Expanded diversity among Californian *Borrelia* isolates and description of *Borrelia bissettii* sp. nov. (formerly *Borrelia* group DN127). *J Clin Microbiol,* 36(12)**,** 3497-504.

PROCTOR, H. & OWENS, I. I. 2000. Mites and birds: diversity, parasitism and coevolution. *Trends Ecol Evol,* 15(9)**,** 358-364.

QIU, W. G., BOSLER, E. M., CAMPBELL, J. R., UGINE, G. D., WANG, I. N., LUFT, B. J. & DYKHUIZEN, D. E. 1997. A population genetic study of *Borrelia burgdorferi* sensu stricto from eastern Long Island, New York, suggested frequency-dependent selection, gene flow and host adaptation. *Hereditas,* 127(3)**,** 203-16.

QIU, W. G., DYKHUIZEN, D. E., ACOSTA, M. S. & LUFT, B. J. 2002. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. *Genetics,* 160(3)**,** 833-49.

QIU, W. G., SCHUTZER, S. E., BRUNO, J. F., ATTIE, O., XU, Y., DUNN, J. J., FRASER, C. M., CASJENS, S. R. & LUFT, B. J. 2004. Genetic exchange and

plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A,* 101(39)**,** 14150-5.
RANDOLPH, S. E. & CRAINE, N. G. 1995. General framework for comparative quantitative studies on transmission of tick-borne diseases using Lyme borreliosis in Europe as an example. *J Med Entomol,* 32(6)**,** 765-77.
RANDOLPH, S. E., GREEN, R. M., HOODLESS, A. N. & PEACEY, M. F. 2002. An empirical quantitative framework for the seasonal population dynamics of the tick *Ixodes ricinus*. *Int J Parasitol,* 32(8)**,** 979-89.
RANDOLPH, S. E., MIKLISOVA, D., LYSY, J., ROGERS, D. J. & LABUDA, M. 1999. Incidence from coincidence: patterns of tick infestations on rodents facilitate transmission of tick-borne encephalitis virus. *Parasitology,* 118 (Pt 2)177-86.
RANDOLPH, S. E. & STOREY, K. 1999. Impact of microclimate on immature tick-rodent host associations (Acari: Ixodidae): Implications for parasite transmission. *Journal of Medical Entomology,* 36(6)**,** 741-748.
RAUTER, C. & HARTUNG, T. 2005. Prevalence of *Borrelia burgdorferi* sensu lato genospecies in *Ixodes ricinus* ticks in Europe: a metaanalysis. *Appl Environ Microbiol,* 71(11)**,** 7203-16.
RICHTER, D., POSTIC, D., SERTOUR, N., LIVEY, I., MATUSCHKA, F. R. & BARANTON, G. 2006. Delineation of *Borrelia burgdorferi* sensu lato species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp. nov. *Int J Syst Evol Microbiol,* 56(Pt 4)**,** 873-81.
RIJPKEMA, S. G., MOLKENBOER, M. J., SCHOULS, L. M., JONGEJAN, F. & SCHELLEKENS, J. F. 1995. Simultaneous detection and genotyping of three genomic groups of *Borrelia burgdorferi* sensu lato in Dutch *Ixodes ricinus* ticks by characterization of the amplified intergenic spacer region between 5S and 23S rRNA genes. *J Clin Microbiol,* 33(12)**,** 3091-5.
RIJPKEMA, S. G., TAZELAAR, D. J., MOLKENBOER, M. J., NOORDHOEK, G. T., PLANTINGA, G., SCHOULS, L. M. & SCHELLEKENS, J. F. 1997. Detection of *Borrelia* afzelii, *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* and group VS116 by PCR in skin biopsies of patients with erythema migrans and acrodermatitis chronica atrophicans. *Clin Microbiol Infect,* 3(1)**,** 109-116.
ROBINSON, D. F. & FOULDS, L. R. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences,* 53131-147.
ROGERS, A. R. & HARPENDING, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol,* 9(3)**,** 552-69.
ROSA, P. A., SCHWAN, T. & HOGAN, D. 1992. Recombination between genes encoding major outer surface proteins A and B of *Borrelia burgdorferi*. *Mol Microbiol,* 6(20)**,** 3031-40.
RUDENKO, N., GOLOVCHENKO, M., GRUBHOFFER, L. & OLIVER, J. H., JR. 2009a. *Borrelia* carolinensis sp. nov., a new (14th) member of the *Borrelia burgdorferi* Sensu Lato complex from the southeastern region of the United States. *J Clin Microbiol,* 47(1)**,** 134-41.
RUDENKO, N., GOLOVCHENKO, M., LIN, T., GAO, L., GRUBHOFFER, L. & OLIVER, J. H., JR. 2009b. Delineation of a new species of the *Borrelia burgdorferi* Sensu Lato Complex, *Borrelia* americana sp. nov. *J Clin Microbiol,* 47(12)**,** 3875-80.
RUDENKO, N., GOLOVCHENKO, M., MOKRACEK, A., PISKUNOVA, N., RUZEK, D., MALLATOVA, N. & GRUBHOFFER, L. 2008. Detection of *Borrelia bissettii* in cardiac valve tissue of a patient with endocarditis and aortic valve stenosis in the Czech Republic. *J Clin Microbiol,* 46(10)**,** 3540-3.
RUDENKO, N., GOLOVCHENKO, M., RUZEK, D., PISKUNOVA, N., MALLATOVA, N. & GRUBHOFFER, L. 2009c. Molecular detection of *Borrelia* bissettii DNA in

serum samples from patients in the Czech Republic with suspected borreliosis. *FEMS Microbiol Lett,* 292(2)**,** 274-81.
RUZIC-SABLJIC, E., LOTRIC-FURLAN, S., MARASPIN, V., CIMPERMAN, J., PLETERSKI-RIGLER, D. & STRLE, F. 2001. Analysis of *Borrelia burgdorferi* sensu lato isolated from cerebrospinal fluid. *APMIS,* 109(10)**,** 707-13.
SADZIENE, A., WILSKE, B., FERDOWS, M. S. & BARBOUR, A. G. 1993. The cryptic ospC gene of *Borrelia burgdorferi* B31 is located on a circular plasmid. *Infect Immun,* 61(5)**,** 2192-5.
SCHARLEMANN, J. P., JOHNSON, P. J., SMITH, A. A., MACDONALD, D. W. & RANDOLPH, S. E. 2008. Trends in Ixodid tick abundance and distribution in Great Britain. *Med Vet Entomol,* 22(3)**,** 238-47.
SCHMIDT, H. A., STRIMMER, K., VINGRON, M. & VON HAESELER, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics,* 18(3)**,** 502-4.
SCHULZE, L., JORDAN, R. A., DOLAN, M. C., DIETRICH, G., HEALY, S. P. & PIESMAN, J. 2008. Ability of 4-poster passive topical treatment devices for deer to sustain low population levels of Ixodes scapularis (Acari: Ixodidae) after integrated tick management in a residential landscape. *J Med Entomol,* 45(5)**,** 899-904.
SCHWAN, T. G. 2003. Temporal regulation of outer surface proteins of the Lyme-disease spirochaete *Borrelia burgdorferi*. *Biochem Soc Trans,* 31(Pt 1)**,** 108-12.
SCHWAN, T. G., BURGDORFER, W. & GARON, C. F. 1988. Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi*, as a result of in vitro cultivation. *Infect Immun,* 56(8)**,** 1831-6.
SCHWAN, T. G. & HINNEBUSCH, B. J. 1998. Bloodstream- versus tick-associated variants of a relapsing fever bacterium. *Science,* 280(5371)**,** 1938-40.
SCHWAN, T. G., SCHRUMPF, M. E., KARSTENS, R. H., CLOVER, J. R., WONG, J., DAUGHERTY, M., STRUTHERS, M. & ROSA, P. A. 1993. Distribution and molecular analysis of Lyme disease spirochetes, *Borrelia burgdorferi*, isolated from ticks throughout California. *J Clin Microbiol,* 31(12)**,** 3096-108.
SCHWARTZ, J. J., GAZUMYAN, A. & SCHWARTZ, I. 1992. rRNA gene organization in the Lyme disease spirochete, *Borrelia burgdorferi*. *J Bacteriol,* 174(11)**,** 3757-65.
SCHWEIZER, M., EXCOFFIER, L. & HECKEL, G. 2007. Fine-scale genetic structure and dispersal in the common vole (*Microtus arvalis*). *Molecular Ecology,* 162463-2473.
SEARLE, J. B., KOTLIK, P., RAMBAU, R. V., MARKOVA, S., HERMAN, J. S. & MCDEVITT, A. D. 2009. The Celtic fringe of Britain: insights from small mammal phylogeography. *Proc Biol Sci,* 276(1677)**,** 4287-94.
SHEAVES, B. J. & BROWN, R. W. 1995. Densities of *Ixodes ricinus* ticks (Acari: Ixodidae) on moorland vegetation communities in the UK. *Exp Appl Acarol,* 19(9)**,** 489-97.
SMITH, R. P., JR., MUZAFFAR, S. B., LAVERS, J., LACOMBE, E. H., CAHILL, B. K., LUBELCZYK, C. B., KINSLER, A., MATHERS, A. J. & RAND, P. W. 2006. *Borrelia garinii* in seabird ticks (Ixodes uriae), Atlantic Coast, North America. *Emerg Infect Dis,* 12(12)**,** 1909-12.
SPIELMAN, A., WILSON, M. L., LEVINE, J. F. & PIESMAN, J. 1985. Ecology of *Ixodes dammini*-borne human babesiosis and Lyme disease. *Annu Rev Entomol,* 30439-60.
STACKEBRANDT, E. & EBERS, J. 2006. Taxonomic parameters reisited: Tarnished gold standards. *Microbiology Today,* 33152-155.
STAFFORD, K. C., 3RD, DENICOLA, A. J. & KILPATRICK, H. J. 2003. Reduced abundance of *Ixodes scapularis* (Acari: Ixodidae) and the tick parasitoid

Ixodiphagus hookeri (Hymenoptera: Encyrtidae) with reduction of white-tailed deer. *J Med Entomol,* 40(5)**,** 642-52.
STALHAMMAR-CARLEMALM, M., JENNY, E., GERN, L., AESCHLIMANN, A. & MEYER, J. 1990. Plasmid analysis and restriction fragment length polymorphisms of chromosomal DNA allow a distinction between *Borrelia burgdorferi* strains. *Zentralbl Bakteriol,* 274(1)**,** 28-39.
STEERE, A. C., COBURN, J. & GLICKSTEIN, L. 2004. The emergence of Lyme disease. *J Clin Invest,* 113(8)**,** 1093-101.
STEVENSON, B. & BARTHOLD, S. W. 1994. Expression and sequence of outer surface protein C among North American isolates of *Borrelia burgdorferi*. *FEMS Microbiol Lett,* 124(3)**,** 367-72.
STEVENSON, B. & MILLER, J. C. 2003. Intra- and interbacterial genetic exchange of Lyme disease spirochete erp genes generates sequence identity amidst diversity. *J Mol Evol,* 57(3)**,** 309-24.
STRLE, F., NELSON, J. A., RUZIC-SABLJIC, E., CIMPERMAN, J., MARASPIN, V., LOTRIC-FURLAN, S., CHENG, Y., PICKEN, M. M., TRENHOLME, G. M. & PICKEN, R. N. 1996. European Lyme Borreliosis: 231 culture-confirmed cases involving patients with erythema migrans. *Clin Infect Dis,* 23(1)**,** 61-5.
SWANSON, K. I. & NORRIS, D. E. 2008. Presence of multiple variants of *Borrelia burgdorferi* in the natural reservoir *Peromyscus leucopus* throughout a transmission season. *Vector Borne Zoonotic Dis,* 8(3)**,** 397-405.
TABERLET, P. & BOUVET, J. 1994. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear Ursus arctos in Europe. *Proc Biol Sci,* 255(1344)**,** 195-200.
TABERLET, P., FUMAGALLI, L. & HAUSSER, J. 1994. Chromosomal versus mitochondria1 DNA evolution: tracking thc cvolutiotrary history of the southwestern European populations of the Sow arunm group (Mammalia, Insectivora). *Evolution,* 48623-636.
TAKANO, A., GOKA, K., UNE, Y., SHIMADA, Y., FUJITA, H., SHIINO, T., WATANABE, H. & KAWABATA, H. 2010. Isolation and characterization of a novel *Borrelia* group of tick-borne *Borrelia*e from imported reptiles and their associated ticks. *Environ Microbiol,* 12(1)**,** 134-46.
TALLEKLINT, L. & JAENSON, T. G. 1997. Infestation of mammals by *Ixodes ricinus* ticks (Acari: Ixodidae) in south-central Sweden. *Exp Appl Acarol,* 21(12)**,** 755-71.
TAO, N., RICHARDSON, R., BRUNO, W. & KUIKEN, C. 2009. *FindModel* [Online]. Los Alamos: HIV Sequence Database. Available: http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html [Accessed 5/02/2010 2010].
TARAGEL'OVA, V., KOCI, J., HANINCOVA, K., KURTENBACH, K., DERDAKOVA, M., OGDEN, N. H., LITERAK, I., KOCIANOVA, E. & LABUDA, M. 2008. Blackbirds and song thrushes constitute a key reservoir of *Borrelia garinii*, the causative agent of Borreliosis in Central Europe. *Appl Environ Microbiol,* 74(4)**,** 1289-93.
TEREKHOVA, D., IYER, R., WORMSER, G. P. & SCHWARTZ, I. 2006. Comparative genome hybridization reveals substantial variation among clinical isolates of *Borrelia burgdorferi* sensu stricto with different pathogenic properties. *J Bacteriol,* 188(17)**,** 6124-34.
THEISEN, M., FREDERIKSEN, B., LEBECH, A. M., VUUST, J. & HANSEN, K. 1993. Polymorphism in ospC gene of *Borrelia burgdorferi* and immunoreactivity of OspC protein: implications for taxonomy and for use of OspC protein as a diagnostic antigen. *J Clin Microbiol,* 31(10)**,** 2570-6.
URWIN, R. & MAIDEN, M. C. 2003. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol,* 11(10)**,** 479-87.

VAN DAM, A. P. 2002. Diversity of *Ixodes*-borne *Borrelia* species--clinical, pathogenetic, and diagnostic implications and impact on vaccine development. *Vector Borne Zoonotic Dis,* 2(4)**,** 249-54.

VAN DAM, A. P., KUIPER, H., VOS, K., WIDJOJOKUSUMO, A., DE JONGH, B. M., SPANJAARD, L., RAMSELAAR, A. C., KRAMER, M. D. & DANKERT, J. 1993. Different genospecies of *Borrelia burgdorferi* are associated with distinct clinical manifestations of Lyme borreliosis. *Clin Infect Dis,* 17(4)**,** 708-17.

VAN DAM, A. P., OEI, A., JASPARS, R., FIJEN, C., WILSKE, B., SPANJAARD, L. & DANKERT, J. 1997. Complement-mediated serum sensitivity among spirochetes that cause Lyme disease. *Infect Immun,* 65(4)**,** 1228-36.

VAN OVERBEEK, L., GASSNER, F., VAN DER PLAS, C. L., KASTELEIN, P., NUNES-DA ROCHA, U. & TAKKEN, W. 2008. Diversity of *Ixodes ricinus* tick-associated bacterial communities from different forests. *FEMS Microbiol Ecol,* 66(1)**,** 72-84.

VITORINO, L. R., MARGOS, G., FEIL, E. J., COLLARES-PEREIRA, M., ZE-ZE, L. & KURTENBACH, K. 2008. Fine-scale phylogeographic structure of *Borrelia* lusitaniae revealed by multilocus sequence typing. *PLoS One,* 3(12)**,** e4002.

VOLLMER, S. A., BORMANE, A., DINNIS, R. E., SEELIG, F., DOBSON, A. D. M., AANENSEN, D. M., JAMES, M. C., DONAGHY, M., RANDOLPH, S. E., FEIL, E., KURTENBACH, K. & MARGOS, G. 2010. Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. *Environ Microbiol,* in press(in press).

VOS, M. & DIDELOT, X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME J,* 3(2)**,** 199-208.

WANG, G., VAN DAM, A. P. & DANKERT, J. 1999a. Evidence for frequent *OspC* gene transfer between *Borrelia valaisiana* sp. nov. and other Lyme disease spirochetes. *FEMS Microbiol Lett,* 177(2)**,** 289-96.

WANG, G., VAN DAM, A. P. & DANKERT, J. 2000. Two distinct ospA genes among *Borrelia valaisiana* strains. *Res Microbiol,* 151(5)**,** 325-31.

WANG, G., VAN DAM, A. P., LE FLECHE, A., POSTIC, D., PETER, O., BARANTON, G., DE BOER, R., SPANJAARD, L. & DANKERT, J. 1997. Genetic and phenotypic analysis of *Borrelia valaisiana* sp. nov. (*Borrelia* genomic groups VS116 and M19). *Int J Syst Bacteriol,* 47(4)**,** 926-32.

WANG, G., VAN DAM, A. P., SCHWARTZ, I. & DANKERT, J. 1999b. Molecular typing of *Borrelia burgdorferi* sensu lato: taxonomic, epidemiological, and clinical implications. *Clin Microbiol Rev,* 12(4)**,** 633-53.

WANG, G., VAN DAM, A. P., SPANJAARD, L. & DANKERT, J. 1998. Molecular typing of *Borrelia burgdorferi* sensu lato by randomly amplified polymorphic DNA fingerprinting analysis. *J Clin Microbiol,* 36(3)**,** 768-76.

WANG, I. N., DYKHUIZEN, D. E., QIU, W., DUNN, J. J., BOSLER, E. M. & LUFT, B. J. 1999c. Genetic diversity of ospC in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics,* 151(1)**,** 15-30.

WARD, A. I. 2005. Expanding ranges of wild and feral deer in Great Britain. *Mammal Rev,* 35(2)**,** 165-173.

WAYNE, L. G., BRENNER, R. R., COLWELL, P. A. D., GRIMONT, O., KANDLER, M. I., MOORE, L. H., MOORE, W. E. C., MURRAY, G. E., STACKEBRANDT, E., STARR, M. P. & TRUPER, H. G. 1987. Report of the Ad Hoc Committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol.,* 37463-464.

WHITLOCK, M. C. 1996. The Red Queen Beats the Jack-Of-All-Trades: The Limitations on the Evolution of Phenotypic Plasticity and Niche Breadth. *The American Naturalist,* 14865.

WIENECKE, R., ZOCHLING, N., NEUBERT, U., SCHLUPEN, E. M., MEURER, M. & VOLKENANDT, M. 1994. Molecular subtyping of *Borrelia burgdorferi* in erythema migrans and acrodermatitis chronica atrophicans. *J Invest Dermatol,* 103(1)**,** 19-22.

WILEY, E. O. 1988. Vicariance biogeography. *Annual Review of Ecological Systematics,* 19513-542.

WILSKE, B. 2003. Diagnosis of Lyme Borreliosis in europe. *Vector Borne Zoonotic Dis,* 3(4)**,** 215-27.

WILSKE, B., BUSCH, U., EIFFERT, H., FINGERLE, V., PFISTER, H. W., ROSSLER, D. & PREAC-MURSIC, V. 1996a. Diversity of OspA and OspC among cerebrospinal fluid isolates of *Borrelia burgdorferi* sensu lato from patients with neuroborreliosis in Germany. *Med Microbiol Immunol,* 184(4)**,** 195-201.

WILSKE, B., BUSCH, U., FINGERLE, V., JAURIS-HEIPKE, S., PREAC MURSIC, V., ROSSLER, D. & WILL, G. 1996b. Immunological and molecular variability of OspA and OspC. Implications for *Borrelia* vaccine development. *Infection,* 24(2)**,** 208-12.

WILSKE, B., JAURIS-HEIPKE, S., LOBENTANZER, R., PRADEL, I., PREAC-MURSIC, V., ROSSLER, D., SOUTSCHEK, E. & JOHNSON, R. C. 1995. Phenotypic analysis of outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato by monoclonal antibodies: relationship to genospecies and OspA serotype. *J Clin Microbiol,* 33(1)**,** 103-9.

WILSKE, B., PREAC-MURSIC, V., GOBEL, U. B., GRAF, B., JAURIS, S., SOUTSCHEK, E., SCHWAB, E. & ZUMSTEIN, G. 1993. An OspA serotyping system for *Borrelia burgdorferi* based on reactivity with monoclonal antibodies and OspA sequence analysis. *J Clin Microbiol,* 31(2)**,** 340-50.

WILSKE, B., PREAC-MURSIC, V., SCHIERZ, G., KUHBECK, R., BARBOUR, A. G. & KRAMER, M. 1988. Antigenic variability of *Borrelia burgdorferi*. *Ann N Y Acad Sci,* 539126-43.

WILSKE, B. & SCHRIEFER, M. 2003. *Manual of Clinical Microbiology,* Washington, DC, ASM Press.

WILSKE, B., ZOLLER, V., BRADE, V., EIFFERT, H., GOBEL, U. B. & STANEK, G. 2000. *Quality Standards for the Microbiological Diagnosis of Infectious Diseases* [Online]. Munich: NRZ fur *Borrelia*n. Available: http://pollux.mpk.med.uni-muenchen.de/alpha1/nrz-*Borrelia*/miq-lyme/frame-miq-lyme.html [Accessed 11 August 2010 2010].

WILSON, M. L. 1986. Reduced abundance of adult Ixodes dammini (Acari: Ixodidae) following destruction of vegetation. *J Econ Entomol,* 79(3)**,** 693-6.

WILSON, M. L., ALDER, G. H. & SPEIELMAN, A. 1985. Correlation between abundance of deer and that of the deer tick, *Ixodes dommini* (Acari: Ixodidae. *Annual Entomological Society of America,* 78172-176.

WILSON, M. L., LEVINE, J. F. & SPIELMAN, A. 1984. Effect of deer reduction on abundance of the deer tick (Ixodes dammini). *Yale J Biol Med,* 57(4)**,** 697-705.

WOOLHOUSE, M. E., TAYLOR, L. H. & HAYDON, D. T. 2001. Population biology of multihost pathogens. *Science,* 292(5519)**,** 1109-12.

WYWIAL, E., HAVEN, J., CASJENS, S. R., HERNANDEZ, Y. A., SINGH, S., MONGODIN, E. F., FRASER-LIGGETT, C. M., LUFT, B. J., SCHUTZER, S. E. & QIU, W. G. 2009. Fast, adaptive evolution at a bacterial host-resistance locus: the PFam54 gene array in *Borrelia burgdorferi*. *Gene,* 445(1-2)**,** 26-37.

XU, G., FANG, Q. Q., KEIRANS, J. E. & DURDEN, L. A. 2003. Molecular phylogenetic analyses indicate that the *Ixodes ricinus* complex is a paraphyletic group. *J Parasitol,* 89(3)**,** 452-7.

YANAGIHARA, Y. & MASUZAWA, T. 1997. Lyme disease (Lyme borreliosis). *FEMS Immunol Med Microbiol,* 18(4)**,** 249-61.

YANG, X. F., PAL, U., ALANI, S. M., FIKRIG, E. & NORGARD, M. V. 2004. Essential role for OspA/B in the life cycle of the Lyme disease spirochete. *J Exp Med,* 199(5)**,** 641-8.

ZHIOUA, E., AESCHLIMANN, A. & GERN, L. 1994. Infection of field-collected *Ixodes ricinus* (Acari: Ixodidae) larvae with *Borrelia burgdorferi* in Switzerland. *J Med Entomol,* 31(5)**,** 763-6.

Appendix 1

Table A1.1 Strains used in this study and their sequence type (ST) and geographic source.

Strain	Species	ST	Collection Site	Region	Country	Year	Decimal degrees
61527BT	*B. afzelii*	164	WH	Bath	England	2006	51.375717 -2.33625
6311BT	*B. afzelii*	164	WH	Bath	England	2006	51.375717 -2.33625
6515BT	*B. afzelii*	164	WH	Bath	England	2006	51.375717 -2.33625
70587B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
71086B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
71186B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
71980B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
74486B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
74586B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
76694B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
77680B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
78094B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
79080B	*B. afzelii*	164	WH	Bath	England	2007	51.375717 -2.33625
85620B	*B. afzelii*	240	Exmoor	Somerset	England	2008	51.183836 -3.571426
80225B	*B. afzelii*	241	NF II	Hampsh-ire	England	2008	50.79636 -1.646018
80710B	*B. afzelii*	250	RW	Bath	England	2008	51.363588 -2.336633
80810B	*B. afzelii*	250	RW	Bath	England	2008	51.363588 -2.336633
81310B	*B. afzelii*	250	RW	Bath	England	2008	51.363588 -2.336633
81410B	*B. afzelii*	250	RW	Bath	England	2008	51.363588 -2.336633
82516B	*B. afzelii*	250	WF	Wiltshire	England	2008	51.3705 -2.293733
82020B	*B. afzelii*	265	Exmoor	Somerset	England	2008	51.183836 -3.571426
83519B	*B. afzelii*	265	Exmoor	Somerset	England	2008	51.183836 -3.571426
84019B	*B. afzelii*	265	Exmoor	Somerset	England	2008	51.183836 -3.571426
IBS-11	*B. afzelii*	72	n/a	Alsace	France	n/a	n/a
IBS-12	*B. afzelii*	73	n/a	Alsace	France	n/a	n/a
IPT118	*B. afzelii*	74	n/a	Auvergne	France	n/a	45.738083 3.049611
IBS-13	*B. afzelii*	75	n/a	Alsace	France	n/a	n/a
IPT154	*B. afzelii*	76	n/a	Limousin	France	n/a	46.15975 1.842111
IPT109	*B. afzelii*	77	n/a	Alsace	France	n/a	48.028611 7.198194
IPT122	*B. afzelii*	77	n/a	Auvergne	France	n/a	45.752333 3.02375
IPT164	*B. afzelii*	77	n/a	Auvergne	France	n/a	45.752333 3.02375
IPT179	*B. afzelii*	77	n/a	Auvergne	France	n/a	45.760389 2.985972
IPT138	*B. afzelii*	78	n/a	Alsace	France	n/a	48.019056 7.132528
IPT152	*B. afzelii*	79	n/a	Limousin	France	n/a	46.15975 1.842111
IPT142	*B. afzelii*	80	n/a	Alsace	France	n/a	47.920306 7.206722
IPT110	*B. afzelii*	81	n/a	Alsace	France	n/a	47.92275 7.153639
80401G	*B. afzelii*	80	Sb	Bonn	Germany	2008	50.673917 7.247672
91506G	*B. afzelii*	204	LM	Sauerland	Germany	2009	51.114327 8.052077
81705G	*B. afzelii*	242	Kottenf-orst	Bonn	Germany	2008	50.707574 7.085452
80504G	*B. afzelii*	249	Kottenf-orst	Bonn	Germany	2008	50.707574 7.085452
91305G	*B. afzelii*	254	LM	Sauerland	Germany	2009	51.114327 8.052077
91206G	*B. afzelii*	255	LM	Sauerland	Germany	2009	51.114327 8.052077
91006G	*B. afzelii*	256	LM	Sauerland	Germany	2009	51.114327 8.052077
92906G	*B. afzelii*	257	LM	Sauerland	Germany	2009	51.114327 8.052077
90208G	*B. afzelii*	258	Wichmar	Saxony	Germany	2009	51.035417 11.68396
92405G	*B. afzelii*	259	LM	Sauerland	Germany	2009	51.114327 8.052077
90607G	*B. afzelii*	260	Wichmar	Saxony	Germany	2009	51.035417 11.68396
91505G	*B. afzelii*	261	LM	Sauerland	Germany	2009	51.114327 8.052077

63203L	*B. afzelii*	165	Babite	Riga	Latvia	2006	56.83278 23.790211
64409L	*B. afzelii*	165	Babite	Riga	Latvia	2006	56.83278 23.790211
64224L	*B. afzelii*	165	Jaunc-iems	Riga	Latvia	2006	57.053433 24.1561
75021L	*B. afzelii*	165	Jaunc-iems	Riga	Latvia	2007	57.053433 24.1561
71424L	*B. afzelii*	166	Jaunc-iems	Riga	Latvia	2007	57.053433 24.1561
61218L	*B. afzelii*	166	Kemeri	Riga	Latvia	2006	56.934667 23.4841
60618L	*B. afzelii*	167	Kemeri	Riga	Latvia	2006	56.934667 23.4841
70515L	*B. afzelii*	168	Kemeri	Riga	Latvia	2007	56.934667 23.4841
60724L	*B. afzelii*	169	Jaunc-iems	Riga	Latvia	2006	57.053433 24.1561
63321L	*B. afzelii*	170	Jaunc-iems	Riga	Latvia	2006	57.053433 24.1561
63721L	*B. afzelii*	170	Jaunc-iems	Riga	Latvia	2006	57.053433 24.1561
75918L	*B. afzelii*	170	Kemeri	Riga	Latvia	2007	56.934667 23.4841
74618L	*B. afzelii*	171	Kemeri	Riga	Latvia	2007	56.934667 23.4841
71618L	*B. afzelii*	204	Kemeri	Riga	Latvia	2007	56.934667 23.4841
63521L	*B. afzelii*	215	Jaunc-iems	Riga	Latvia	2006	57.053433 24.1561
61918L	*B. afzelii*	215	Kemeri	Riga	Latvia	2006	56.934667 23.4841
62218L	*B. afzelii*	215	Kemeri	Riga	Latvia	2006	56.934667 23.4841
72721L	*B. afzelii*	216	Jaunc-iems	Riga	Latvia	2007	57.053433 24.1561
72915L	*B. afzelii*	217	Kemeri	Riga	Latvia	2007	56.934667 23.4841
63424L	*B. afzelii*	219	Jaunc-iems	Riga	Latvia	2006	57.053433 24.1561
71924L	*B. afzelii*	220	Jaunc-iems	Riga	Latvia	2007	57.053433 24.1561
703105B	*B. afzelii*	263	Inverness	Morayshire	Scotland	2007	57.616052 -3.716633
710118B	*B. afzelii*	263	Inverness	Morayshire	Scotland	2007	57.616052 -3.716633
702123B	*B. afzelii*	264	Inverness	Morayshire	Scotland	2007	57.616052 -3.716633
71509L	*B. burgdorferi*	161	Babite	Riga	Latvia	2007	56.83278 23.790211
74306L	*B. burgdorferi*	161	Babite	Riga	Latvia	2007	56.83278 23.790211
72506L	*B. burgdorferi*	162	Babite	Riga	Latvia	2007	56.83278 23.790211
IPT190	*B. burgdorferi*	20	n/a	Normandie	France	n/a	48.58125 0.499111
IPT191	*B. burgdorfe*	20	n/a	Normandie	France	n/a	48.814306 1.056806
IPT2	*B. burgdorferi*	20	n/a	Alsace	France	n/a	48.019056 7.132528
IPT23	*B. burgdorferi*	20	n/a	Alsace	France	n/a	48.019444 7.167444
IPT69	*B. burgdorferi*	20	n/a	Alsace	France	n/a	48.022444 7.12675
IPT19	*B. burgdorferi*	21	n/a	Alsace	France	n/a	47.920306 7.206722
IPT135	*B. burgdorferi*	22	n/a	Auvergne	France	n/a	45.760389 2.985972
IPT137	*B. burgdorferi*	23	n/a	Alsace	France	n/a	47.92275 7.153639
IPT39	*B. burgdorferi*	24	n/a	Alsace	France	n/a	48.056 7.167444
IPT58	*B. burgdorferi*	24	n/a	Alsace	France	n/a	48.022444 7.12675
IPT193	*B. burgdorferi*	25	n/a	Normandie	France	n/a	48.542694 0.368639
IPT198	*B. burgdorferi*	25	n/a	Normandie	France	n/a	48.542694 0.368639
84214B	*B. garinii*	82	BHW	Bath	England	2008	51.385717 -2.32015

80235B	*B. garinii*	82	RP	London	England	2008	51.441042 -0.270377
71180B	*B. garinii*	82	WH	Bath	England	2007	51.375717 -2.33625
679BT	*B. garinii*	86	BHW	Bath	England	2006	51.385717 -2.32015
70879B	*B. garinii*	86	BHW	Bath	England	2007	51.385717 -2.32015
81640B	*B. garinii*	86	LD	Cumbria	England	2008	n/a
70207B	*B. garinii*	87	BHW	Bath	England	2007	51.385717 -2.32015
70276B	*B. garinii*	87	BHW	Bath	England	2007	51.385717 -2.32015
73294B	*B. garinii*	87	WH	Bath	England	2007	51.375717 -2.33625
71057B	*B. garinii*	87	Winsley	Wiltshire	England	2007	51.3461 -2.29605
82514B	*B. garinii*	88	BHW	Bath	England	2008	51.385717 -2.32015
81329B	*B. garinii*	88	NF II	Hampshire	England	2008	50.79636 -1.646018
80602B	*B. garinii*	93	NF I	Hampshire	England	2008	50.83325 -1.57225
70242B	*B. garinii*	93	Winsley	Wiltshire	England	2007	51.3461 -2.29605
64418BT	*B. garinii*	163	BHW	Bath	England	2006	51.385717 -2.32015
71063B	*B. garinii*	163	BHW	Bath	England	2007	51.385717 -2.32015
71077B	*B. garinii*	163	BHW	Bath	England	2007	51.385717 -2.32015
71287B	*B. garinii*	163	WH	Bath	England	2007	51.375717 -2.33625
70268B	*B. garinii*	172	Campus	Bath	England	2007	51.38075 -2.327267
73580B	*B. garinii*	172	WH	Bath	England	2007	51.375717 -2.33625
81222B	*B. garinii*	173	Exmoor	Somerset	England	2008	51.183836 -3.571426
70260B	*B. garinii*	173	Winsley	Wiltshire	England	2007	51.3461 -2.29605
70576B	*B. garinii*	174	BHW	Bath	England	2007	51.385717 -2.32015
78180B	*B. garinii*	175	WH	Bath	England	2007	51.375717 -2.33625
6128BT	*B. garinii*	176	WH	Bath	England	2006	51.375717 -2.33625
61710BT	*B. garinii*	179	BHW	Bath	England	2006	51.385717 -2.32015
61030BT	*B. garinii*	183	WF	Wiltshire	England	2006	51.3705 -2.293733
61210BT	*B. garinii*	187	BHW	Bath	England	2006	51.385717 -2.32015
83215B	*B. garinii*	187	BHW	Bath	England	2008	51.385717 -2.32015
85013B	*B. garinii*	187	BHW	Bath	England	2008	51.385717 -2.32015
82310B	*B. garinii*	187	RW	Bath	England	2008	51.363588 -2.336633
70250B	*B. garinii*	187	Campus	Bath	England	2007	51.38075 -2.327267
70489B	*B. garinii*	187	Campus	Bath	England	2007	51.38075 -2.327267
60809BT	*B. garinii*	190	BHW	Bath	England	2006	51.385717 -2.32015
6725BT	*B. garinii*	190	BHW	Bath	England	2006	51.385717 -2.32015
73163B	*B. garinii*	190	BHW	Bath	England	2007	51.385717 -2.32015
72680B	*B. garinii*	190	WH	Bath	England	2007	51.375717 -2.33625
6910BT	*B. garinii*	191	BHW	Bath	England	2006	51.385717 -2.32015
70531B	*B. garinii*	207	BHW	Bath	England	2007	51.385717 -2.32015
73802B	*B. garinii*	207	BHW	Bath	England	2007	51.385717 -2.32015
84715B	*B. garinii*	207	BHW	Bath	England	2008	51.385717 -2.32015
70643B	*B. garinii*	207	WF	Wiltshire	England	2007	51.3705 -2.293733
74694B	*B. garinii*	207	WH	Bath	England	2007	51.375717 -2.33625
80914B	*B. garinii*	243	BHW	Bath	England	2008	51.385717 -2.32015
81315B	*B. garinii*	243	BHW	Bath	England	2008	51.385717 -2.32015
83013B	*B. garinii*	243	BHW	Bath	England	2008	51.385717 -2.32015
82239B	*B. garinii*	244	RW	Bath	England	2008	51.363588 -2.336633
83439B	*B. garinii*	244	RW	Bath	England	2008	51.363588 -2.336633
83016B	*B. garinii*	244	WF	Wiltshire	England	2008	51.3705 -2.293733
83113B	*B. garinii*	245	BHW	Bath	England	2008	51.385717 -2.32015
80215B	*B. garinii*	246	BHW	Bath	England	2008	51.385717 -2.32015
80813B	*B. garinii*	246	BHW	Bath	England	2008	51.385717 -2.32015
82814B	*B. garinii*	246	BHW	Bath	England	2008	51.385717 -2.32015
85513B	*B. garinii*	246	BHW	Bath	England	2008	51.385717 -2.32015

80330B	*B. garinii*	246	RP	London	England	2008	51.441042 -0.270377
IPT157	*B. garinii*	83	n/a	Limousin	France	n/a	46.030806 1.285278
IPT156	*B. garinii*	86	n/a	Auvergne	France	n/a	45.752333 3.02375
IPT167	*B. garinii*	86	n/a	Limousin	France	n/a	46.030806 1.285278
IPT168	*B. garinii*	86	n/a	Limousin	France	n/a	46.030806 1.285278
IPT178	*B. garinii*	86	n/a	Auvergne	France	n/a	45.738083 3.049611
IPT189	*B. garinii*	86	n/a	Norman-die	France	n/a	48.58125 0.499111
IPT28	*B. garinii*	86	n/a	Alsace	France	n/a	48.019056 7.132528
IPT140	*B. garinii*	87	n/a	Alsace	France	n/a	48.028611 7.198194
IPT114	*B. garinii*	88	n/a	Alsace	France	n/a	47.92275 7.153639
IPT158	*B. garinii*	88	n/a	Limousin	France	n/a	46.007972 1.29325
IPT195	*B. garinii*	88	n/a	Norman-die	France	n/a	48.58125 0.499111
IPT130	*B. garinii*	89	n/a	Alsace	France	n/a	47.907583 7.158611
IPT139	*B. garinii*	90	n/a	Alsace	France	n/a	47.920306 7.206722
IPT165	*B. garinii*	91	n/a	Auvergne	France	n/a	45.752333 3.02375
IPT169	*B. garinii*	92	n/a	Auvergne	France	n/a	45.752333 3.02375
IPT171	*B. garinii*	93	n/a	Auvergne	France	n/a	45.760389 2.985972
IPT172	*B. garinii*	94	n/a	Auvergne	France	n/a	45.760389 2.985972
80505G	*B. garinii*	84	Kottenf-orst	Bonn	Germany	2008	50.707574 7.085452
91102G	*B. garinii*	94	Kottenf-orst	Bonn	Germany	2009	50.707574 7.085452
91001G	*B. garinii*	179	Kottenf-orst	Bonn	Germany	2009	50.707574 7.085452
81201G	*B. garinii*	180	Sb	Bonn	Germany	2008	50.673917 7.247672
90104G	*B. garinii*	187	Kottenf-orst	Bonn	Germany	2009	50.707574 7.085452
91806G	*B. garinii*	187	LM	Sauerland	Germany	2009	51.114327 8.052077
81001G	*B. garinii*	251	Sb	Bonn	Germany	2008	50.673917 7.247672
82805G	*B. garinii*	252	Kottenf-orst	Bonn	Germany	2008	50.707574 7.085452
80201G	*B. garinii*	253	Sb	Bonn	Germany	2008	50.673917 7.247672
91701G	*B. garinii*	262	Kottenf-orst	Bonn	Germany	2009	50.707574 7.085452
61006L	*B. garinii*	86	Babite	Riga	Latvia	2006	56.83744 23.787186
65006L	*B. garinii*	86	Babite	Riga	Latvia	2006	56.83278 23.790211
70903L	*B. garinii*	86	Babite	Riga	Latvia	2007	56.83278 23.790211
74215L	*B. garinii*	86	Kemeri	Riga	Latvia	2007	56.934667 23.4841
73606L	*B. garinii*	89	Babite	Riga	Latvia	2007	56.83278 23.790211
75309L	*B. garinii*	89	Babite	Riga	Latvia	2007	56.83278 23.790211
71609L	*B. garinii*	90	Babite	Riga	Latvia	2007	56.83278 23.790211
61406L	*B. garinii*	163	Babite	Riga	Latvia	2006	56.83278 23.790211
70809L	*B. garinii*	163	Babite	Riga	Latvia	2007	56.83278 23.790211
61009L	*B. garinii*	177	Babite	Riga	Latvia	2006	56.83744 23.787186
76418L	*B. garinii*	178	Kemeri	Riga	Latvia	2007	56.934667 23.4841
62306L	*B. garinii*	180	Babite	Riga	Latvia	2006	56.83278 23.790211
62506L	*B. garinii*	180	Babite	Riga	Latvia	2006	56.83278 23.790211
62118L	*B. garinii*	180	Kemeri	Riga	Latvia	2006	56.934667 23.4841
74503L	*B. garinii*	181	Babite	Riga	Latvia	2007	56.83278 23.790211
74109L	*B. garinii*	182	Babite	Riga	Latvia	2007	56.83278 23.790211
70406L	*B. garinii*	184	Babite	Riga	Latvia	2007	56.83278 23.790211
63809L	*B. garinii*	185	Babite	Riga	Latvia	2006	56.83278 23.790211
610112L	*B. garinii*	186	Babite	Riga	Latvia	2006	56.83744 23.787186
73912L	*B. garinii*	187	Babite	Riga	Latvia	2007	56.83278 23.790211
62303L	*B. garinii*	188	Babite	Riga	Latvia	2006	56.83278 23.790211
62309L	*B. garinii*	188	Babite	Riga	Latvia	2006	56.83278 23.790211

61209L	*B. garinii*	189	Babite	Riga	Latvia	2006	56.83278 23.790211
62909L	*B. garinii*	190	Babite	Riga	Latvia	2006	56.83278 23.790211
72409L	*B. garinii*	190	Babite	Riga	Latvia	2007	56.83278 23.790211
69212L	*B. garinii*	193	Babite	Riga	Latvia	2006	56.83278 23.790211
74415L	*B. garinii*	207	Kemeri	Riga	Latvia	2007	56.934667 23.4841
75803L	*B. garinii*	208	Babite	Riga	Latvia	2007	56.83278 23.790211
74403L	*B. garinii*	209	Babite	Riga	Latvia	2007	56.83278 23.790211
66612L	*B. garinii*	214	Babite	Riga	Latvia	2006	56.83278 23.790211
71412L	*B. lusitania*	218	Babite	Riga	Latvia	2007	56.83278 23.790211
80105G	*B. spielmanii*	239	Kottenf-orst	Bonn	Germany	2008	50.707574 7.085452
70323B	*B. valaisiana*	96	HH	Hamp-shire	England	2007	51.3104 -0.906683
80310B	*B. valaisiana*	96	RW	Bath	England	2008	51.363588 -2.336633
71987B	*B. valaisiana*	96	WH	Bath	England	2007	51.375717 -2.33625
702100B	*B. valaisiana*	97	BHW	Bath	England	2007	51.385717 -2.32015
712100B	*B. valaisiana*	97	BHW	Bath	England	2007	51.385717 -2.32015
81313B	*B. valaisiana*	97	BHW	Bath	England	2008	51.385717 -2.32015
85213B	*B. valaisiana*	97	BHW	Bath	England	2008	51.385717 -2.32015
70155B	*B. valaisiana*	97	WF	Wiltshire	England	2007	51.3705 -2.293733
71185B	*B. valaisiana*	97	WH	Bath	England	2007	51.375717 -2.33625
81102B	*B. valaisiana*	99	NF I	Hamp-shire	England	2008	50.83325 -1.57225
6136BT	*B. valaisiana*	99	WH	Bath	England	2006	51.375717 -2.33625
70183B	*B. valaisiana*	102	BHW	Bath	England	2007	51.385717 -2.32015
80515B	*B. valaisiana*	102	BHW	Bath	England	2008	51.385717 -2.32015
70278B	*B. valaisiana*	102	Campus	Bath	England	2007	51.38075 -2.327267
70479B	*B. valaisiana*	103	BHW	Bath	England	2007	51.385717 -2.32015
70512B	*B. valaisiana*	103	BHW	Bath	England	2007	51.385717 -2.32015
71587B	*B. valaisiana*	192	WH	Bath	England	2007	51.375717 -2.33625
61104BT	*B. valaisiana*	195	WH	Bath	England	2006	51.375717 -2.33625
61214BT	*B. valaisiana*	196	WH	Bath	England	2006	51.375717 -2.33625
70134B	*B. valaisiana*	198	BHW	Bath	England	2007	51.385717 -2.32015
81020B	*B. valaisiana*	199	Exmoor	Somerset	England	2008	51.183836 -3.571426
61230BT	*B. valaisiana*	199	WF	Wiltshire	England	2006	51.3705 -2.293733
621BT	*B. valaisiana*	200	ThW	Somerset	England	2006	50.982263 -3.001413
82240B	*B. valaisiana*	201	LD	Cumbria	England	2008	n/a
81702B	*B. valaisiana*	201	NF I	Hamp-shire	England	2008	50.83325 -1.57225
81510B	*B. valaisiana*	201	RW	Bath	England	2008	51.363588 -2.336633
633BT	*B. valaisiana*	201	WH	Bath	England	2006	51.375717 -2.33625
6930BT	*B. valaisiana*	205	WF	Wiltshire	England	2006	51.3705 -2.293733
70655B	*B. valaisiana*	205	WF	Wiltshire	England	2007	51.3705 -2.293733
70113B	*B. valaisiana*	210	BHW	Bath	England	2007	51.385717 -2.32015
74094B	*B. valaisiana*	211	WH	Bath	England	2007	51.375717 -2.33625
75980B	*B. valaisiana*	211	WH	Bath	England	2007	51.375717 -2.33625
78380B	*B. valaisiana*	211	WH	Bath	England	2007	51.375717 -2.33625
70865B	*B. valaisiana*	212	Campus	Bath	England	2007	51.38075 -2.327267
81124B	*B. valaisiana*	247	Exmoor	Somerset	England	2008	51.183836 -3.571426
80319B	*B. valaisiana*	248	Exmoor	Somerset	England	2008	51.183836 -3.571426
IPT102	*B. valaisiana*	96	n/a	Auvergne	France	n/a	45.738083 3.049611
IPT163	*B. valaisiana*	96	n/a	Auvergne	France	n/a	45.738083 3.049611
IPT174	*B. valaisiana*	96	n/a	Auvergne	France	n/a	45.760389 2.985972
IPT29	*B. valaisiana*	96	n/a	Meuse	France	n/a	48.979167 5.329167
IPT85	*B. valaisiana*	96	n/a	Alsace	France	n/a	47.92275 7.153639
IPT177	*B. valaisiana*	97	n/a	Limousin	France	n/a	45.998333 1.415444

IPT188	*B. valaisiana*	97	n/a	Norm-andie	France	n/a	48.610917 0.459028
IPT31	*B. valaisiana*	97	n/a	Meuse	France	n/a	49.022222 5.25
IPT111	*B. valaisiana*	98	n/a	Alsace	France	n/a	47.920306 7.206722
IPT33	*B. valaisiana*	98	n/a	Meuse	France	n/a	49.075 5.261111
IPT47	*B. valaisiana*	99	n/a	Alsace	France	n/a	48.022444 7.12675
IPT121	*B. valaisiana*	100	n/a	Alsace	France	n/a	47.92275 7.153639
IPT144	*B. valaisiana*	101	n/a	Limousin	France	n/a	46.15975 1.842111
IPT184	*B. valaisiana*	102	n/a	Limousin	France	n/a	46.15975 1.842111
IPT187	*B. valaisiana*	102	n/a	Limousin	France	n/a	46.15975 1.842111
IPT186	*B. valaisiana*	103	n/a	Limousin	France	n/a	46.15975 1.842111
61306L	*B. valaisiana*	96	Babite	Riga	Latvia	2006	56.83278 23.790211
72009L	*B. valaisiana*	96	Babite	Riga	Latvia	2007	56.83278 23.790211
61809L	*B. valaisiana*	97	Babite	Riga	Latvia	2006	56.83278 23.790211
63503L	*B. valaisiana*	97	Babite	Riga	Latvia	2006	56.83278 23.790211
73009L	*B. valaisiana*	97	Babite	Riga	Latvia	2007	56.83278 23.790211
74003L	*B. valaisiana*	97	Babite	Riga	Latvia	2007	56.83278 23.790211
63603L	*B. valaisiana*	102	Babite	Riga	Latvia	2006	56.83278 23.790211
72509L	*B. valaisiana*	197	Babite	Riga	Latvia	2007	56.83278 23.790211
72609L	*B. valaisiana*	199	Babite	Riga	Latvia	2007	56.83278 23.790211
73815L	*B. valaisiana*	199	Kemeri	Riga	Latvia	2007	56.934667 23.4841
73915L	*B. valaisiana*	199	Kemeri	Riga	Latvia	2007	56.934667 23.4841
62103L	*B. valaisiana*	201	Babite	Riga	Latvia	2006	56.83278 23.790211
61618L	*B. valaisiana*	202	Kemeri	Riga	Latvia	2006	56.934667 23.4841
71512L	*B. valaisiana*	203	Babite	Riga	Latvia	2007	56.83278 23.790211
75618L	*B. valaisiana*	203	Kemeri	Riga	Latvia	2007	56.934667 23.4841
72209L	*B. valaisiana*	206	Babite	Riga	Latvia	2007	56.83278 23.790211
72403L	*B. valaisiana*	211	Babite	Riga	Latvia	2007	56.83278 23.790211
71515L	*B. valaisiana*	211	Kemeri	Riga	Latvia	2007	56.934667 23.4841
610412L	*B. valaisiana*	212	Babite	Riga	Latvia	2006	56.83744 23.787186
63909L	*B. valaisiana*	212	Babite	Riga	Latvia	2006	56.83278 23.790211
68412L	*B. valaisiana*	212	Babite	Riga	Latvia	2006	56.83278 23.790211
73409L	*B. valaisiana*	213	Babite	Riga	Latvia	2007	56.83278 23.790211
90101B	*B. valaisiana*	196	RD	The Gower	Wales	2009	n/a

Abbreviations for sites are as follows Bathampton woods (BHW), Hazeley Heath (HH), New Forest I (NF I), New Forest II (NF II), Lake District (LD), Lennestadt-Meggen (LM), Rainbow Woods (RW), Rhossili Downs (RD), Richmond Park (RP), Siebenbeirge (Sb), Thurlbear woods (ThW), Warleigh Forrest (WF).

Table A1.2 Nymphal and adult tick densities from collections at English sites from 2006 to 2009 including data on humidity and temperature.

Site	Year	Date	Month	Season	Mean Herbage	Humidity (%)	Temp (°C)	Area (m^2)	Nymphs	Adults	Nymphs/m	Adults/m
Bathampton	2006	27/4/06	Apr	Spring	20	58	15	45	3	0	0.07	0.00
Bathampton	2006	28/4/06	Apr	Spring	30	65	14	68	51	3	0.75	0.04
Bathampton	2006	30/4/06	Apr	Spring	15	60	13	56	7	1	0.13	0.02
Bathampton	2006	1/5/06	May	Spring	18	70	10	64	14	2	0.22	0.03
Bathampton	2006	11/5/06	May	Spring	32	58	22	46	38	2	0.83	0.04
Bathampton	2006	12/8/06	Aug	Summer	n/a	66	16	87	33	2	0.38	0.02
Bathampton	2006	27/9/06	Sep	Autumn	n/a	65	19	24	3	0	0.13	0.00
Brendon Forest	2006	30/4/06	Apr	Spring	57	n/a	12	101	71	7	0.70	0.07
Warleigh	2006	27/9/06	Sep	Autumn	n/a	75	17	58	6	0	0.10	0.00
Warleigh	2006	28/9/06	Sep	Autumn	n/a	65	19	74	8	0	0.11	0.00
Widcombe	2006	22/4/06	Apr	Spring	0	60	17	80	19	0	0.24	0.00
Widcombe	2006	28/4/06	Apr	Spring	3	58	15	75	17	0	0.23	0.00
Widcombe	2006	29/4/06	Apr	Spring	5	65	14	127	28	0	0.22	0.00
Widcombe	2006	1/5/06	May	Spring	21	60	13	103	18	0	0.17	0.00
Widcombe	2006	2/5/06	May	Spring	17	64	12	67	20	1	0.30	0.01
Widcombe	2006	11/5/06	May	spring	25	58	17	25	6	4	0.24	0.16
Widcombe	2006	12/5/06	May	Spring	33	58	22	104	42	0	0.40	0.00
Widcombe	2006	12/8/06	Aug	Summer	n/a	65	17	44	7	0	0.16	0.00
Bathampton	2007	28/3/07	Mar	Spring	7	54	16	60	23	2	0.38	0.03
Bathampton	2007	3/4/07	Apr	Spring	n/a	47	17	66	31	4	0.47	0.06
Bathampton	2007	14/4/07	Apr	Spring	36	70	19	98	33	5	0.34	0.05
Bathampton	2007	28/4/07	Apr	Spring	8	75	23	42	35	2	0.83	0.05
Bathampton	2007	5/5/07	May	Spring	51	70	17	87	48	2	0.55	0.02
Bathampton	2007	26/5/07	May	Spring	91	79	16	81	32	12	0.40	0.15
Bathampton	2007	2/6/07	Jun	Summer	50	80	15	30	1	1	0.03	0.03
Bathampton	2007	1/8/07	Aug	Summer	95	43	22	50	8	0	0.16	0.00
Bathampton	2007	3/8/07	Aug	Summer	67.5	61	18	63	7	0	0.11	0.00
Bathampton	2007	25/8/07	Aug	Summer	14	60	20	47.5	4	0	0.08	0.00
Bathampton	2007	1/9/07	Sep	Autumn	20	67	18	87.5	19	1	0.22	0.01

Bathampton	2007	15/9/07	Sep	Autumn	n/a	78	18	122	13	3	0.11	0.02
Bathampton	2007	20/10/07	Oct	Autumn	26	65	13	100	6	0	0.06	0.00
Bathampton	2007	3/11/07	Nov	Autumn	8	71	16	130	27	1	0.21	0.01
Campus	2007	28/4/07	Apr	Spring	95	75	23	30.5	2	3	0.07	0.10
Campus	2007	5/5/07	May	Spring	95	70	17	53.5	2	11	0.04	0.21
Campus	2007	19/5/07	May	Spring	100	92	14	26	6	2	0.23	0.08
Campus	2007	2/6/07	Jun	Summer	90	80	15	32.5	0	6	0.00	0.18
Campus	2007	1/8/07	Aug	Summer	95	43	22	40	0	3	0.00	0.08
Campus	2007	3/8/07	Aug	Summer	92	61	18	80	1	11	0.01	0.14
Campus	2007	1/9/07	Sep	Autumn	88	67	18	81	1	6	0.01	0.07
Campus	2007	6/9/07	Sep	Autumn	67.5	61	22	106	2	2	0.02	0.02
Campus	2007	14/9/07	Sep	Autumn	89	95	13	90	5	6	0.06	0.07
Campus	2007	2/10/07	Oct	Autumn	85	65	12	100	17	1	0.17	0.01
Campus	2007	7/10/07	Oct	Autumn	66	75	12	113	15	1	0.13	0.01
Campus	2007	3/11/07	Nov	Autumn	76	71	16	100	15	1	0.15	0.01
Hazeley Heath	2007	9/4/07	Apr	Spring	40	n/a	16	68.5	23	0	0.34	0.00
Rowas	2007	22/10/07	Oct	Autumn	5	n/a	12	150	38	0	0.25	0.00
Rowas	2007	2/11/07	Nov	Autumn	5	n/a	12	98	36	3	0.37	0.03
Rowas	2007	5/11/07	Nov	Autumn	13	n/a	13	85	19	1	0.22	0.01
Warleigh	2007	24/4/07	Apr	Spring	88	60	14	65.5	28	0	0.43	0.00
Warleigh	2007	1/5/07	May	Spring	69	70	22	87.5	30	1	0.34	0.01
Warleigh	2007	17-Sep	Sep	Autumn	n/a	50	19	100	6	0	0.06	0.00
Warleigh	2007	23/10/07	Oct	Autumn	50	65	12	155	13	1	0.08	0.01
Warleigh	2007	5/11/07	Nov	Autumn	43	90	12	157	10	0	0.06	0.00
Widcombe	2007	11/8/07	Aug	Summer	7	61	18	94.5	22	4	0.23	0.04
Widcombe	2007	9/9/07	Sep	Autumn	25	55	23	55	17	2	0.31	0.04
Widcombe	2007	14/9/07	Sep	Autumn	6	63	19	106	23	3	0.22	0.03
Widcombe	2007	20/10/07	Oct	Autumn	4	66	12	207	30	1	0.14	0.00
Bathampton	2008	5/3/08	Mar	Spring	2.5	n/a	n/a	54	34	2	0.63	0.04
Bathampton	2008	21/5/08	May	Spring	26	n/a	n/a	115	101	12	0.88	0.10
Bathampton	2008	19/6/08	Jun	Summer	16	n/a	n/a	50	55	1	1.10	0.02
Bathampton	2008	22/8/08	Aug	Summer	12	n/a	n/a	51	15	0	0.29	0.00

Campus	2008	22/8/07	Aug	Summer	60	n/a	n/a	115	6	2	0.05	0.02
Campus	2008	8/2/08	Feb	Winter	11	65	12	112	12	1	0.11	0.01
Campus	2008	5/3/08	Mar	Spring	46	n/a	n/a	51	2	0	0.04	0.00
Campus	2008	21/5/08	May	Spring	89	n/a	n/a	70	8	10	0.11	0.14
Rainbow woods	2008	9/4/08	Apr	Spring	11	n/a	n/a	164	13	0	0.08	0.00
Rainbow woods	2008	10/6/08	Jun	Summer	53	n/a	20	148	9	3	0.06	0.02
Rainbow woods	2008	22/8/08	Aug	Summer	46	n/a	n/a	163	17	7	0.10	0.04
Rowas	2008	6/5/08	May	Spring	22	n/a	n/a	173.5	28	6	0.16	0.03
Rowas	2008	23/7/08	Jul	Summer	16	n/a	22	145	73	9	0.50	0.06
Warleigh	2008	24/7/08	July	Summer	58	n/a	17	150	66	9	0.44	0.06
Rhossili downs	2009	5/4/09	Apr	Spring	65	n/a	20	130	62	9	0.48	0.07
Rhossili downs	2009	5/4/09	Apr	Spring	100	n/a	20	100	20	0	0.20	0.00
Rosetherne	2009	9/5/09	May	Spring	80	n/a	13	104	4	0	0.04	0.00

Appendix 2

Table A2.1 Primer sets used for gene amplification including single loci and the eight housekeeping genes. Table also shows the position within the gene or region and the expected size of product.

Primer and target gene	Sequence 5' to 3'	Position	Maximum fragment length (bp)
IGS (5S-23S)			
Outer Reverse	ACCATAGACTCTTATTACTTTGAC	469-446	
Outer Forward	TAAGCTGACTAATACTAATTACCC	92-115	
Inner Reverse	ACCATAGACTCTTATTACTTTGACCA	469-444	
Inner Forward	GAGAGTAGGTTATTGCCAGGG	243-363	225
ospA			
Outer Reverse	GACCTTAAAGGAACTTCTGATAA	356-334	
Outer Forward	GTATTGTTGTACTGTAATTGT	874-894	
Inner Reverse	ATGGATCTGGAGTACTTGAA	381-362	
Inner Forward	CTTAAAGTAACAGTTCCTTCT	693-713	350
New outer forward	GGGRATAGGTCTAATATTAG	18-38	
New outer reverse	GCGTTTTTAAGTTCATCAAG	832-855	
New inner forward	TAATATTAGCCTTAATAGCATG	37-55	
New inner reverse	CGTATTTTTGTACWGTWATTG	829-845	808
clpA semi-nested			
Outer Forward	AAAGATAGATTTCTTCCAGAC	1237	
Reverse	GAATTTCATCTATTAAAAGCTTTC	2218	
Inner Forward	AAAGCTTTTGATATTTTAGATG	1258	706
clpX			
Outer Forward	GCTGCAGAGATGAATGTGCC	391-410	
Outer Reverse	GATTGATTTCATATAACTCTTTTG	1250-1273	
Inner Forward	AATGTGCCATTTGCAATAGC	403	
Inner Reverse	TTAAGAAGACCCTCTAAAATAG	1124	721
***nifS* semi-nested**			
Forward	ATGGATTTCAAACAAATAAAAAG	1-23	
Outer Reverse	GTTGGAGCAAGCATTTTATG	719	
Inner Reverse	5´-GATATTATTGAATTTCTTTTAAG-3´	1027-1049	685
pepX			
Outer Forward	ACAGAGACTTAAGCTTAGCAG	362	
Outer Reverse	GTTCCAATGTCAATAGTTTC	1153-1172	
Inner Forward	TTATTCCAAACCTTGCAATCC	449-469	
Inner Reverse	TGTGCCTGAAGGAACATTTG	1097-1115	666
pyrG			
Outer Forward	GATTGCAAGTTCTGAGAATA	391-410	
Outer Reverse	CAAACATTACGAGCAAATTC	1190	
Inner Forward	GATATGGAAAATATTTTATTTATTG	448	
Inner Reverse	AAACCAAGACAAATTCCAAG	1135-1154	687
recG			
Outer Forward	CCCTTGTTGCCTTGCTTTC	890-908	
Outer Reverse	GAAAGTCCAAAACGCTCAG	1694	
Inner Forward	CTTTAATTGAAGCTGGATATC	917	
Inner Reverse	CAAGTTGCATTTGGACAATC	1639-1658	741
***rplB* semi-nested**			
Outer Forward	TGGGTATTAAGACTTATAAGC	2-22	
Inner Forward	CGCTATAAGACGACTTTATC	40-59	
Reverse	GCTGTCCCCAAGGAGACA	743-760	720
uvrA			
Outer Forward	GAAATTTTAAAGGAAATTAAAAGTAG	1408	
Outer Reverse	CAAGGAACAAAAACATCTGG	2299-2318	
Inner Forward	GCTTAAATTTTTAATTGATGTTGG	1434-1457	
Inner Reverse	CCTATTGGTTTTTGATTTATTTG	2111	677

Appendix 3

Table A3.1 Allele codes assigned to the intergenic spacer region (5S-23S) sequence fragments and the strains that were found to contain the alleles.

Species	Allele name	Strain	Sequence
B. valaisiana	V1	73915L	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACTTTATTTAATAAAACATTCAAA AACATAACATTAAAAAAATATAAAAAATAAATTTAATTTAA ATTACAAAATAAAAACCCTCAAAACC
	V2	60201BT	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAACATTAAAAAAATATAAAAAATAAATTTAATTTAA ATACAAAATAAAAACCCTAAAACAAA
	V3	610412L 68412L 63909L 70865B 73409L	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAACATTATAAAAATATAAAAAATAAATTTAATTTAA ATACAAAATAAAAACCCTCAATAAC
	V4	60930BT	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAACATTAAAAAAATATAAAAAATAAATTTAATTTAA ATACAAAATAAAAACCCTCATAAAC
	V5	61214BT	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAAACATTAAAAAAATATAAAAAATAAATTTAATTTA AATTACAAAATAAAAACCCTCATAAAC
	V6	61230BT	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAACATTAAAAAAATATAAAAAATAAATTTAATTTAA ATTACAAAATAAAAACCCTCATAAAC
	V7	72509L Bvalaisiana VS116 70323B IPT29Bv IPT85Bv IPT102Bv IPT144Bv IPT163Bv IPT174Bv 71987B 72009L 61306L	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAAACATTAAAAAAATATAAAAAATAAATTTAATTTA AATTACAAAATAAAAACCCTCAATAAC
	V8	IPT33Bv IPT111	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAAACATTAAAAAAATAAATTTAATTTAAATTACAAA ATAAAAACCCTCAATAAC
	V9	71515L 61618L 62103L	

		63603L	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAACATTAAAAAAATATAAAAAATAAATTTAATTTAA ATACAAAATAAAAACCCTCAATAAC
		63503L	
		61809L	
		702100B	
		74094B	
		IPT47	
		IPT121	
		IPT184	
		IPT187	
		61306BT	
		60303BT	
		75618L	
		70278B	
		72209L	
	V10	712100B	TTATCAACATAAAATAATATATATCTTTTTTCAATCCATTCA ATATCTATTTTATTTTTTACATTATTTAATAAAACATTCAAA AACATAACATTAAAAAAATATAAAAAATAAATTTAATTTAA ATTACAAAATAAAAACCCTCAATAAC
		71587B	
		IPT31	
		IPT177	
		IPT186	
		IPT188	
B. garinii	G1	62309L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATATTTTATTTTTTATTTTTTAATTTTATATTATTTAA TTTTTATTCAAAAAATATAAACATCAAAAAACATAAAAAAT AAAATCAATTTTAAATATAAAATAAAAACCCTCAATAAC
	G2	60725BT	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAAATATCTAAAAACATAAAAAATAAAATCAATTTTAA ATATAAAATAAAAACCCTCANAAAC
	G3	74109L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAAACATCTAAAAACATAAAAAATAAAATCAATTTTAA ATATAAAATAAAAACCCTCAATAAC
		74503L	
		62506L	
		72506LB	
		62118L	
		62306L	
		IPT114	
		IPT158	
		IPT195	
	G4	73912L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATATTTTATTTTTTATATTATTTAATAAAACATTCAA ATAATATAAACATCTAAAAACATAAAAAATAAAATCAATTT TAAATATAAAATAAAAACCCTCAATAAC
		70489B	
		70250B	
		IPT172	
		IPT171	
		IPT139	
	G5	IPT169	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATATTTTATTTTTTATATTATTTAATAAAACATTCAA ATAATATAAACATCTAAAAACATAAAAAATAAAATCAATTT TAAATATAAAAAAAAACCCTCAATAAC
	G6	60809BT	TTATACAACATAAATAATATATATCTTTTTTAATCCANTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAAATATCTAAAAACATAAAAAATAAAATCAATTTTAA ATATAAAATAAAAACCCTCAATAAC

	G7	73163B	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAAATATCTAAAAACATAAAAAATAAAATCAATTTTAA ATATAAAATAAAAACCCTCAATAAC
		72409L	
		IPT165	
		IPT178	
		610112L	
	G8	74415L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAACATCTAAAAACATAAAAAATAAAATCAATAATAAA AACAAAAACCCTCAACAAC
	G9	71180B	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAACATCTAAAAACATAAAAAATAAAATCAATTTTAAA TATAAAACAAAAACCCTCAATAAC
	G10	*B. garinii* 20047	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTCAAAA ACATAACATCTAAAAACATAAAAAATAAAATCAATTTTAAA TATAAAACAAAAACCCTCAACAAC
		74694B	
		IPT157	
	G11	70809L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATATTTTATTTTTTATATTATTCAATAAAACATTCAA AAACATAACATCTAAAAACATAAAAAATAAAATCAATTTTA AATATAAAATAAAAACCCTCAATAAC
		61406L	
		71287B	
		64418BT	
	G12	61006L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTCAATAAAACATTCAAAA ACATAACATCTAAAAACATAAAAAATAAAATCAATTTTAAA TATAAAATAAAAACCCTCAATAAC
		66612L	
		61009L	
		65006L	
		73294B	
		60709BT	
		IPT189	
		IPT168	
		IPT167	
		IPT156	
		IPT140	
		IPT28	
	G13	61030BT	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTAAAAA CATAACATCTAAAAACATAAAAAATAAAATCAATTTTAAAT ATAAAATAAAAACCCTCANAAAC
	G14	63809L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATTTTATTTTTTATATTATTTAATAAAACATTAAAAA CATAACATCTAAAAACATAAAAAATAAAATCAATTTTAAAT ATAAAATAAAAACCCTCAATAAC
		IPT130	
		75803L	
B. afzelii	A1	63424L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA TATATATATTATTTTTTATATTATTTAATAAAACATTCAAAT AATATAAACATTTAAAAAATAAATTCAAATTTAAATATAAA ATAAAAACCCTCAATAAC
		63321L	
		63721L	
		IPT154B	
	A2	61918L	TTATACAACATAAAATAATATATATCTTTTTTAATCCATTCA ATATATATATTATTTTTTATATTATTTAATAAAACATTCAAA TAATATAAACATTTAAAAAATAAATTTAATTTAAATATAAAA TAAAAACCCTCAATAAC
		62218L	
	A3	63521L	TTATACAACATAAAATAATATATATCTTTTTTAATCCATTCA ATATATATATTATTTTTTATATTATTTAATAAAACATTCAAA TAATATAAACATTTAAAAAATAAATTCAATTTAAATATAAAA TAAAAACCCTCAATAAC
	A4	63203L	TCATACAACATAAATAATATATATCTTTTTTAATCCATTCAA

		64224L	TATATATATTATTTTTTATATTATTTAATTTTTATTCAAATAA
		64409L	TATAAACATTTAAAAAATAAATTCAATTTAAATATAAAATAA AAACCCTCAATAAC
	A5	IPT110B	
		IPT142B	
		B. afzelii VS461	
		71924L	
		60618L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA
		61218L	TATATATATTATTTTTTATATTATTTAATTTTTATTCAAATAA TATAAACATTTAAAAAATAAATTCAAATTTAAATATAAAATA
		IPT109B	AAAACCCTCAATAAC
		IPT122B	
		IPT138B	
		IPT164B	
		IPT179B	
		60724L	
	A6	*B. afzelii* Pko	
		71924L	
		60618	
		61218L	TTATACAACATAAATAATATATATCTTTTTTAATCCATTCAA
		IPT109	TATATATATTATTTTTTATATTATTTAATTTTTATTCAAATAA TATAAACATTTAAAAAATAAATTCAATTTAAATATAAAATAA
		IPT122	AAACCCTCAATAAC
		IPT138	
		IPT164	
		IPT179	
	A7	60724L	TTATACAACATAAAATAATATATATCTTTTTTAATCCATTCA ATATATATATTATTTTTTATATTATTTAATTTTTATTCAAATA
		61527BT	ATATAAACATTTAAAAAATAAATTTAATTTAAATATAAAATA AAAACCCTCATAAAC
	A8	60311BT	
		77680B	
		74486B	TTATACAACATAAAATAATATATATCTTTTTTAATCCATTCA
		79080B	ATATATATATTATTTTTTATATTATTTAATTTTTATTCAAATA ATATAAACATTAAAAAATAAATTTAATTTAAATATAAAATAA
		71086B	AAACCCTCAATAAC
		71186B	
		74586B	
	A9	70587B	TTATACAACATAAAATAATATATATCTTTTTTAATCCATTCA ATATATATATTATTTTTTATATTATTTAATTTTTATTCAAATA
		78094B	ATATAAACATTATAAACATTTAAAAAATAAATTTAATTTAAA TATAAAATAAAAACCCTCAATAAC
	A10	76694B	TTATACAACATAAAATAATATATATCTTTTTTAATCCATTCA ATATATATATTATTTTTTATATTATTTAATTTTTATTCAAATA ATATAAACATTTAAAAAATAAATTTAATTTAAATATAAAATA AAAACCCTCAATAAC
B. burgdorferi	B1	IPT118B	TTATCAACATAAATAATATATATCTTTTTTAATACATTCAAT ATATATTTCATTTTTTATTTATTTAAATAACACATTCAAAAA
		IPT193B	CACTAATATTTAAAAACCAAAAATAAAATCAAATTAAAACA TAAAAACAAAAACCCTCAATAAC
	B2	IPT198B	TTATACAACATAAAATAATATATATCTTTTTTAATACATTCA
		IPT2B	ATATATATTTTATTTTTTATTTATTTAAACAACACATTCAAA AACACCAATATTTAAAAAACATAAAAATAAAATCAAATTTA
		IPT19B	AATATAAAAATAAAAACCCTCAATAAC
		IPT23B	
		IPT69B	
		IPT135B	

		IPT137B	
		IPT190B	
		IPT191B	
		B. burg. B31	
	B3	IPT39B	TTATACAACATAAAATAATATATATCTTTTTTAATCATTCAA TATATATATTTTATTTTTTATTTATTTAAACAACACATTCAA AAACACCAATATTTAAAAAACATAAAAATAAAATCAAATTT AAATATAAAAATAAAAACCCTCAATAAC
	B4	IPT58B	TTATACAACATAAAATAATATATATCTTTGTTTAATGCATGT CAATATATATATTTTATTTTTTATGTTATTTAAACAACACAT TCAAAAACACCAATATTAAAAAACATAAAAATAAAATCAAA GTTTAAAGTATAAAAATAAAAACCCTGGCAATAAC

Table A3.2 Allele codes assigned to the *ospA* sequence fragments and the strains that were found to contain the alleles.

Species	Allele name	Strain	Sequence
B. valaisiana	Va	V73009L	CTTGGCGAAACCAAACTTGAAACTTTCAAAGA AGATGGAACATTAGTGTCAAGAAAAGTAAATT TCAACGACAAGTCTTTCACAGAAGAAAAATTC AATGAAAAAGGTGAAGTGTCTGAAAAAATACT AACAAGATCAAACGGAACTACACTTGAATACT CACAAATGACAGATGCTGAAAATGCTACAAAA GCAGTAGAAACTGTAAAAAATGGCATTAAGCT TCCAGGAAATCTTGTAGG
	Vb	V70113B_	CTAAGCGAAACCAAACTTGAAACTTTCAAAGA AGATGGTAAAACATTAGTGTCAAGAAAAGTAA ATTTCAAAGACAAGTCTTTCACAGAAGAAAAA TTCAATGAAAAAGGTGAAGTGTCTGAAAAAAT ACTAACAAGAGAAAACGGAACTACACTTGAAT ACTCACAAATGACAGATGCTGAAAATGCTACA AAAGCAGTAGAACTCTAAAAAATGGCATTAAG CTGAAGGAAATCTGTAGG
	Vc	V70278B_	CTAGGCGAAACCAAACTTGAAACTTTCAAAGA AGATGGAACATTAGTGTCAAGAAAAGTAAATT TCAAAGACAAGTCTTTCACAGAAGAAAAATTC AATGAAAAAGGTGAAGTGTCTGAAAAAATACT AACAAGATCAAACGGAACTACACTTGAATACT CACAAATGACAGATGCTGAAAATGCTACAAAA GCAGTAGAACTCTAAAAAATGGCATTAAGCTT CAGGAAATCTGTAGG
	Vd	V70606L	CTAGGCGAAACCAAACTTGAAACTTTCAAAGA AGATGGAACATTAGTGTCAAGAAAAGTAAATT TCAAAGACAAGTCTTTCACAGAAGAAAAATTC AATGAAAAAGGTGAAGTGTCTGAAAAAATACT AACAAGATCAAACGGAACTACACTTGAATACT CACAAATGACAGATGCTGAAAATGCTACAAAA GCAGTAGAAACTCTAAAAAATGGCATTAAGCT TCCAGGAAATCTTGTAGG
		V610412L	
		V610312L	
		IPT188_Bv	
		IPT187_Bv	
		IPT186_Bv	
		IPT184_Bv	
		IPT177_Bv	
		IPT163_Bv	
		IPT144_Bv	
		IPT121_Bv	
		IPT102_Bv	
		IPT85_Bv	
		IPT31_Bv	
		VS116	
		63328BT	
		61230BT	
		60930BT	
		61209L	
		63603L	
		61306L	
		68412L	
		72812L	
		75618L	
		63909L	
		63806L	
		62103L	
		621BT	
		6136BT	
		633BT	
		72609L	
		72209L	
		72009L	

		72509L	
		75980B_	
		IPT29_Bv	
		IPT174_Bv	
	Ve	70323B_	CTAGGCGAAACCAAACTTGAAACTTTCAAAGA
			AGATGGAACATTAGTGTCAAGAAAAGTAAATT
			TCAAAGACAAGTCTTTCACAGAAGAAAAATTC
			AATGAAAAAGGTGAAGTGTCTGAAAAAATACT
			AACAAGATCAAACGGAACTACACTTGAATACT
			CACAAATGACAGATGCTGAAAATGCTACAAAA
			GCAGTAGAACTCTAAAAAATGGCATTAAGCTT
		V70865B_	CCAGGAAATCTGTAGG
	Vf	V61214BT	CTAGGCGAAACCAAACTTGAAACTTTCAAAGA
			AGATGGAACATTAGTGTCAAGAAAAGTGAACA
			TTTCAAAGACAAGTCTTTCACAGAAGAAAAATT
			CAATAGAAAAAGGTGAAGTGTCTGAAAAAATA
			CTAACAAGATCAAACGGAACTACACTTGAATA
			CTCACAAATGACAGATGCTGAAAATGCTACAA
			AAGCAGTAGAAACTCTAAAAAATGGCATTAAG
			CTTCCAGGAAATCTTGTAGG
	Vg	V70512B_	CTAGGCGAAACCAAACTTGAAACTTTCAAAGA
			AGATGGAACATTAGTGTCAAGAAAAGTAAATT
			TCAAAGACAAGTCTTTCACAGAAGAAAAATTC
			AATGAAAAAGGTGAAGTGTCTGAAAAAATACT
			AACAAGATCAAACGGAACTACACTTGAATACT
			CACAAATGACAGATGCTGAAAATGCTACAAAA
			GCAGTAGAACTCTAAAAAATGGCATTAAGCTC
			CAGGAAATCTGTAGG
	Vh	V74094B_	CTAGGCGAAACCAAACTTGAAACTTTCAAAGA
			AGATGGAACATTAGTGTCAAGAAAAGTAAATT
			TCAAAGACAAGTCTTTCACAGAAGAAAAATTC
			AATGAAAAAGGTGAAGTGTCTGAAAAAATACT
			AACAAGATCAAACGGAACTACACTTGAATACT
			CACAAATGACAGATGCTGAAAATGCTACAAAA
			GCAGTAGAAACTCTAAAAAATGGCATTAAGCT
			TCCAGGAAATCTTGTTGG
	Vi	V74003L	NTAGGCGAAACCAAACTTGAAACTTTCAAAGA
		V71512L	AGATGGAACATTAGTGTCAAGAAAAGTAAATT
			TCAAAGACAAGTCTTTCACAGAAGAAAAATTC
			AATGAAAAAGGTGAAGTGTCTGAAAAAATACT
			AACAAGATCAAACGGAACTACACTTGAATACT
			CACAAATGACAGATGCTGAAAATGCTACAAAA
			GCAGTAGAAACTCTAAAAAATGGCATTAAGCT
		V72403L	TCCAGGAAATCTTGTAGG
	Vj	V62303L	CTAGACGAAACCAAACTTGAAACTTTCAAAGA
			AGATGGCACAACATTAGTGTCAAGAAAAGTAA
			CCCTTAAAGACAAGTCATCAACAGAAGAAAAA
			TTTGACGCAAAAGGTGCTGCCGTAACTGAAAA
			TGTAATAACAAGAAAAGACGGAACCAGACTTG
			AATACACAGGAATGAAAAGTGATGGAAGCGG
			AAAAGCTAAAGAAGTTTTAAAAAACTTTGCCCT
			TGATGGAACTTTAGATACTAG
	Vk	V73409L	CTAGACGAAACCAAACTTGAAACTTTCAAAGA
		IPT33_Bv	AGATGGCACAACATTAGTGTCAAGAAAAGTAA
		IPT47_Bv	CCCTTAAAGACAAGTCATCAACAGAAGAAAAA
		IPT111_Bv	TTTGACGCAAAAGGTGCTGCCGTAACTGAAAA
		V61809L	AGTAATAACAAGAAAAGACGGAACCAGACTTG
			AATACACAGGAATGAAAAGTGATGGAAGCGG
			AAAAGCTAAAGAAGTTTTAAAAAACTTTGCCCT
			TGATGGAACTTTAGATACTAG
	Vl	V63012L	CTAGACGAAACCAAACTTGAAACTTTCAAAGA
			AGATGGCACAACATTAGTGTCAAGAAAAGTAA
			CCCTTAAAGACAAGTCATCAACAGAAGAAAAA
			TTTGACGCAAAAGGTGCTNCCGTAACTGAAAA
			AGTAATAACAAGAAAAGACGGAACCAGACTTG

			AATACACAGGAATGAAAAGTGATGGAAGCGG AAAAGCTAAAGAAGTTTTAAAAAACTTTGCCCT TGATGGAACTTTAGATACTAG
B. afzelii	Aa	A61218L	CTAAGTAAAACCACATTCGAACTTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAG TTCTAAAGACAAAACATCAACAGATGAAATGT TCAATGAAAAAGGTGAATTGTCTGCAAAAACC ATGACAAGAGAAAATGGAACCAAACTTGAATA TACAGAAATGAAAAGCGATGGAACCGGAAAA GCTAAAGAAGTTTTAAAAAACTTTACTCTTGAA GGAAAAGTACCTAA
	Ab	A74486B_	CTAAGTAAAACCACATTCGAACTTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAG TTCTAAAGACAAAACATCAACAGATGAAATGT TCAATGAAAAAGGTGAATTGTCTGCAAAAACC ATGACAAGAGAAAATGGAACCAAACTTGAATA TACAGAAATGAAAAGCGATGGAACCGGAAAA GCTAAAGAAGTTTTAAAAAACTTACTCTGAAG GAAAAGTAGCTAA
	Ac	A63721L	CTAAGTAAAACCACATTCGAACTTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAG TTCTAAAGACAAAACATCAACAGATGAAATGT TCAATGAAAAAGGTGAATTGTCTGCAAAAACC ATGACAAGAGAAAATGGAACCAAACTTGAATA TACAGAAATGAAAAGCGATGGAACCGGAAAA GCTAAAGAAGTTTTAAAAAACTTTACTCTTGAA GGAAAAGTAGCTAA
		A74618L	
		A71618L	
		A63203L	
		A63424L	
		VS461	
		Pko	
		IBS11_Ba	
		IBS12_Ba	
		IBS13_Ba	
		IPT109_Ba	
		IPT110_Ba	
		IPT118_Ba	
		IPT122_Ba	
		IPT138_Ba	
		IPT142_Ba	
		IPT152_Ba	
		IPT154_Ba	
		IPT164_Ba	
		IPT179_Ba	
		A76694B_	
		60311BT	
		A2218LT	
		72915L	
		A71924L	
		A75309L	
		A60724L	
		A20105LT	
		A75918L	
	Ad	A74586B_	CTAAGTAAAACCACATTCGAACTTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAG TTCTAAAGACAAAACATCAACAGATGAAATGT TCAATGAAAAAGGTGAATTGTCTGCAAAAACC ATGACAAGAGAAAATGGAACCAAACTTGAATA TACAGAAATGAAAAGCGATGGAACCGGAAAA GCTAAAGAAGTTTTAAAAAACTTTACTCTGAAG GAAAAGTAGCTAA
		78186B_	
		A71186B_	
		A71186B	
		A71986B_	
B. burgdorferi	Ba	IPT193_Bb	CTAGGCCAAACTACACTTGAAGTTTTAAAAGA AGATGGTAAAACATTAGTATCAAGAAAAGTAA CTTCCAAAGACAAGTCATCAACAGAAGAAAAA TTCAACGAAAAAGGTGAATTAGCTGAAAAAAT AATGACAAGAGCAAACGGAACAAGACTTGAAT ACACAGAAATTAAAAGCGATGGATCCGGAAAA
		IPT198_Bb	

			GCTAAAGAAGTTTTAAAAGACTATGTTCTTGAA GGAACTCTAACTGCTGA
	Bb	B31_Bb	CTAGGTCAAACCACACTTGAAGTTTTCAAAGA AGATGGCAAAACACTAGTATCAAAAAAAGTAA CTTCCAAAGACAAGTCATCAACAGAAGAAAAA TTCAATGAAAAAGGTGAAGTATCTGAAAAAAT AATAACAAGAGCAGACGGAACCAGACTTGAAT ACACAGGAATTAAAAGCGATGGATCTGGAAAA GCTAAAGAGGTTTTAAAAGGCTATGTTCTTGA AGGAACTCTAACTGCTGA
		IPT58_Bb	
		B71509L	
		B72506L	
		B74306L	
	Bc	IPT2_Bb	CTAGGTCAAACCACACTTGAAGTTTTCAAAGA AGATGGCAAAACACTAGTATCAAAAAAAGTAA CTTCCAAAGACAAGTCATCAACAGAAGAAAAA TTCAATGAAAAAGGTGAAGTATCTGAAAAAAT AATAACAAGAGCAGACGGAACCAGACTTGAAT ACACAGAAATTAAAAGCGATGGATCTGGAAAA GCTAAAGAGGTTTTAAAAAGCTATGTTCTTGA AGGAACTTTAACTGCTGA
		IPT19_Bb	
		IPT69_Bb	
		IPT137_Bb	
		IPT191_Bb	
	Bd	IPT39_Bb	CTAGGTCAAACCACACTTGAAGTTTTCAAAGA AGATGGCAAAACACTAGTATCAAAAAAAGTAA CTTCCAAAGACAAGTCATCAACAGAAGAAAAA TTCAATGAAAAAGGTGAAGTATCTGAAAAAAT AATAACAAGAGCAGACGGAACCAGACTTGAAT ACACAGAAATTAAAAGCGATGGATCTGGAAAA GCTAAAGAGGTTTTAAAAGGCTATGTTCTTGA AGGAACTTTAACTGCTGA
		IPT135_Bb	
	Be	IPT23_Bb	CTAGGGCAAACCACACTTGAAGTTTTCAAAGA AGATGGCAAAACACTAGTATCAAAAAAAGTAA CTTCCAAAGACAAGTCATCAACAGAAGAAAAA TTCAATGAAAAAGGTGAAGTATCTGAAAAAAT AATAACAAGAGCAGACGGAACCAGACTTGAAT ACACAGAAATTAAAAGCGATGGATCTGGAAAA GCTAAAGAGGTTTTAAAAGGCTATGTTCTTGA AGGAACTTTAACTGCTGA
		IPT190_Bb	
B. garinii	Ga	G70489B	TTAAATCAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAA TTNTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGACATACAAGACGATGGATCCGGAAAC GCTAAAGAAGTTTTAAAAGGCTTTGTTCTTGA AGGAACTCTAATTGCTGA
	Gb	G72409L	TTAAATCAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAA TTCTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACTAGACTTGAATA CACAGAAATAAAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAGGCTTTGCTCTTGAA GGAACTCTAATTGCTGA
	Gc	G74503L	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TTAACGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAAGCTTTACTCTTGAAG GAACTCTAATTGCTGA
	Gd	G74694B_	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TTAACGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAAGCCTTACTCTTGAAG

		GAACTCTAATTGCTGA
Ge	G74403L	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TTAACGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAAGCCTTACTCTTGAAG GAACTNTAACTGCTGA
Gf	G70531B_	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TTAACGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTAAAAAGCCTTACTCTTGAAG GAACTCTAACTGCTGA
Gh	G73802B_ G70875B_	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TTAACGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTAAAAAGCCTTACTCTGAAGG AACTCTAACTGCTGA
Gi	G75803L G74415L	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TTAACGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAAGCCTTACTCTTGAAG GAACTCTAACTGCTGA
Gj	IPT130_Bg G61030BT	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCACAACATTAGTGTCAAGAAAAGTAAA TTCTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAAAAGACGGAACCAGACTTGAATA CACAGACATACAAAGCAATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAGGCCTTACTCTTGAA GGAACTCTAACTGCTGA
Gk	IPT165_Bg G610112L	TTAAATCAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAA TTCTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAGGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGAAATAAAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAGGCTTTGCTCTTGAA GGAACTCTAGCTGATGA
Gl	G62909L G70406L G6910BT IPT178_Bg G60725BT G60809BT	TTAAATCAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAA TTCTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACTAGACTTGAATA CACAGAAATAAAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAGGCTTTGCTCTTGAA GGAACTCTAACTGATGG
Gm	20047_ospAF	TTAAATCAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAA TTCTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAAAAGACGGAACCAGACTTGAATA CACAGACATACAAAACGATGGATCCGGAAAAG

		CTAAAGAAGTTTTAGCAGGCCTTACTCTTGAA GGAACTCTAACTGCTGA
Gn	IPT139_Bg	TTAAATCAAACCACATTTGAAATTTTTAAAGAA
	IPT172_Bg	GATGGCAAAACATTAGTGTCAAGAAAAGTAAA
	G70242B_	TTCTAAAGACAAGTCATCAACAGAAGAAAAAT
	IPT171_Bg	TTAATGATAAAGGTAAATTAAGTGAAAAAGTA
	IPT158_Bg	GTAACAAGAGCAGACGGAACCAGACTTGAATA CACAGAAATACAAAACGATGGATCCGGAAAAG CTAAAGAAGTTTTAAAAGGCCTTACTCTTGAA GGAACTCTAACTAATGA
Go	G73580B_	TTAAATCAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTGTCAAGAAAAGTAAA TTCTAAAGACAAGTCATCAACAGAAGAAAAAT TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGACATACAAGACGATGGATCCGGAAAC GCTAAAGAAGTTTAAAAGGCCTTACTCTTGAA GGAACTCTAACTGCTGA
Gp	IPT157_Bg	TTAAATCAAACCACATTTGAAATTTTCAAAGAA
	IPT169_Bg	GATGGCAAAACATTAGTGTCAAGAAAAGTAAA
	G70250B_	TTCTAAAGACAAGTCATCAACAGAAGAAAAAT
	G61210BT	TTAATGATAAAGGTAAATTAAGTGAAAAAGTA GTAACAAGAGCAAACGGAACCAGACTTGAATA CACAGACATACAAGACGATGGATCCGGAAAC GCTAAAGAAGTTTTAAAAGGCCTTACTCTTGA AGGAACTCTAACTGCTGA
Gq	70903L	NTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAACATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATGGATCCGGAAAA GCTAAAGAAGTTTTAAAAGACTTTACTCTTGAA GGAACTCTAGCTGCTGA
Gr	70879B_	CTAAGTAAAACCACATTTGAAATCTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAATGAAAAGGGTGAAATATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATAAAACCGGAAAAG CTAAAGAAGATTTTAAAAGACTTTACTCTGAAG GAACTCTAGCTGCTGA
Gs	IPT140_Bg	CTAAGTAAAACCACATTTGAAATTTTCAAAGAA
	G62506L	GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAATATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATGGATCCGGAAAA GCTAAAGAAGTTTTAAAAGACTTTACTCTTGAA GGAACTCTAGCTGCTGA
Gt	62306L	CTAAGTAAAACCACATTTGAAATCTTCAAAGAA
	G62118L	GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAATATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATGGATCCGGAAAA GCTAAAGAAGTTTTAAAAGACTTTACTCTTGAA GGAACTCTAGCTGCTGA
Gu	71063B_	CTAAGTAAAACCACATTTGAAATCTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAATATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATAAAACCGGAAAAG

			CTAAAGAAGTTTAAAAGACTTTACTCTGAAGG AACTCTAGCTGCTGA
	Gv	71077B_	CTAAGTAAAACCACATTTGAAATCTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAATATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATAAAACCGGAAAAG CTAAAGAAGTTTTAAAAGACTTTACTCTGAAG GAACTCTAGCTGCTGA
	Gw	70809L G61406L IPT156_Bg G64418BT	CTAAGTAAAACCACATTTGAAATCTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAATATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATAAAACCGGAAAAG CTAAAGAAGTTTTAAAAGACTTTACTCTTGAAG GAACTCTAGCTGCTGA
	Gx	IPT189_Bg IPT167_Bg IPT114_Bg IPT28_Bg IPT195_Bg IPT168_Bg G679BT G61006L G21317LT G61009L G74109L G61710BT G65006L	CTAAGTCAAACTAAATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAC CCTTAAAGACAAGTCATCAACAGAAGAAAAAT TCAACGAAAAGGGTGAAACATCTGAAAAAACA ATAGTAAGAGCAAATGGAACCAGACTTGAATA CACAGACATAAAAAGCGATGGATCCGGAAAA GCTAAAGAAGTTTTAAAAGACTTTACTCTTGAA GGAACTCTAGCTGCTGA
B. bavariensis	Bar	Pbi	CTAAGTAAAACCACATTTGAAATTTTCAAAGAA GATGGCAAAACATTAGTATCAAAAAAAGTAAA TTCTAAAGATAAGTCATCAATAGAAGAAAAATT CAACGCAAAAGGTGAATTATCTGAAAAAACAA TACTAAGAGCAAACGGAACCAGGCTTGAATAC ACAGAAATAAAAAGCGATGGAACCGGAAAAG CTAAAGAAGTTTTAAAAGACTTTGCTCTTGAAG GAACTCTAGCTGCCGA

Appendix 4

Table S1 Sequence types (STs) identified in strains in Chapter 5 spatial structuring study

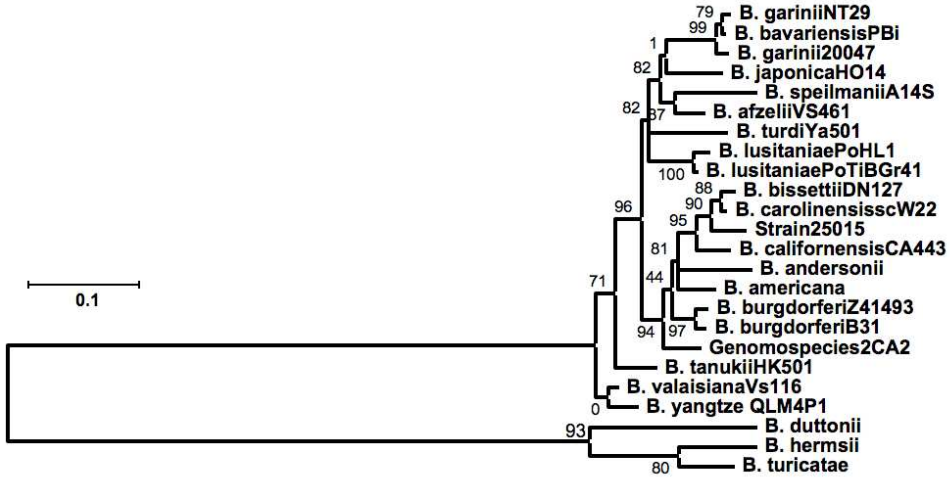
	ST	*clpA*	*clpX*	*nifS*	*pepX*	*pyrB*	*recG*	*rplB*	*uvrA*
B. afzelii	**72**	35	24	24	30	21	27	23	28
	73	37	24	24	31	22	29	23	29
	74	35	24	24	34	21	27	23	28
	75	36	24	23	31	23	30	23	30
	76	36	24	24	31	22	29	23	29
	77	36	24	24	32	21	31	23	28
	78	36	24	23	32	25	32	23	29
	79	36	24	23	32	21	27	23	28
	80	39	24	23	35	26	33	23	29
	81	38	24	25	33	24	29	24	28
	164	35	24	24	32	21	27	23	28
	165	36	24	23	86	22	27	23	28
	166	36	24	23	31	92	27	23	29
	167	36	24	23	30	92	27	23	29
	168	36	24	23	31	85	27	23	29
	169	36	24	23	31	92	89	23	29
	170	37	24	24	31	22	92	23	28
	171	39	24	24	31	22	92	23	28
	204	109	24	23	89	22	27	23	28
	215	51	24	23	86	85	27	23	29
	216	51	24	23	28	85	27	23	29
	217	51	24	23	86	85	91	23	29
	219	102	24	24	31	22	92	23	28
	220	109	24	24	85	90	91	24	29
	240	36	81	80	32	20	95	81	29
	241	111	81	80	32	20	95	81	29
	242	38	24	25	33	93	29	24	28
	249	109	24	24	49	25	27	23	29
	250	113	24	23	32	26	96	23	41
	254	36	24	24	31	22	27	23	29
	255	36	82	24	31	95	27	23	29
	256	37	24	24	101	22	29	23	28
	257	37	24	24	85	22	98	23	29
	258	109	24	23	85	96	27	23	28
	259	109	24	23	89	97	30	23	29
	260	114	24	24	85	22	91	24	29
	261	115	24	24	85	98	91	24	28
	263	37	24	24	31	22	29	23	28
	264	51	24	23	86	92	27	23	28
	265	39	24	23	32	21	29	23	7727
B. burgdorferi	**161**	15	9	12	8	1	17	8	16
	162	15	9	12	8	1	83	8	16
	20	14	1	11	1	1	1	1	10
	21	14	1	11	1	1	10	1	10
	22	14	1	11	1	1	12	1	10
	23	16	1	11	1	1	1	1	10
	24	15	9	12	8	1	11	8	16
	25	13	10	13	10	9	9	7	15
B. garinii	**82**	40	25	26	36	27	34	25	31
	83	40	25	26	42	33	34	25	31
	84	41	26	27	37	28	35	26	32
	86	42	27	29	38	29	36	27	33
	87	42	27	29	38	29	39	27	33
	88	43	28	30	39	30	36	28	34
	89	44	29	31	40	31	37	29	35
	90	45	30	32	41	32	38	30	36
	91	46	31	29	43	34	40	31	37
	92	47	32	33	42	35	41	32	36
	93	45	33	34	36	36	38	30	38
	94	45	30	32	36	32	38	30	36
	163	31	80	78	99	81	39	79	87

	172	40	25	26	36	33	34	25	31
	173	42	27	29	38	29	36	27	82
	174	42	27	29	97	29	36	27	33
	175	42	27	29	38	29	80	27	33
	176	42	27	29	38	29	36	27	79
	177	42	27	29	92	29	36	27	33
	178	42	27	29	38	81	36	27	33
	179	43	28	30	90	82	36	28	34
	180	43	28	30	90	87	36	28	34
	181	43	28	30	39	30	78	28	34
	182	43	28	30	39	88	87	28	34
	183	44	29	76	40	31	37	80	35
	184	44	29	31	40	31	37	80	35
	185	44	29	31	40	31	87	80	77
	186	46	76	73	43	34	40	31	37
	187	47	73	33	42	91	76	32	36
	188	47	32	33	42	91	86	32	36
	189	47	32	33	42	35	88	32	36
	190	48	34	34	44	27	42	33	39
	191	48	79	34	44	27	42	33	39
	193	48	76	29	43	34	42	31	37
	207	95	74	34	96	83	78	77	85
	208	95	29	34	91	89	78	77	85
	209	95	74	34	96	89	78	77	85
	214	99	77	36	91	88	84	75	33
	243	46	76	29	43	34	42	31	37
	244	48	34	34	44	37	42	33	39
	245	99	77	81	91	88	84	82	33
	246	112	80	78	99	81	39	79	87
	251	95	74	34	96	89	78	77	85
	252	47	73	33	42	94	97	32	36
	253	42	27	37	38	29	36	27	33
	262	116	34	34	44	37	42	33	39
B. lusitaniae	**218**	101	21	20	27	86	85	74	81
B. spielmanii	**239**	94	72	79	100	80	94	72	76
B. valaisiana	**96**	49	35	35	45	38	43	35	40
	97	50	36	36	45	38	44	35	40
	98	49	35	35	46	38	43	35	40
	99	49	37	37	45	39	45	36	40
	100	50	38	36	45	38	44	35	40
	101	49	35	38	45	38	43	35	40
	102	50	39	36	45	38	44	35	40
	103	49	36	36	45	38	44	35	40
	192	50	36	36	45	38	44	35	88
	195	49	37	37	45	39	77	36	40
	196	49	35	77	45	38	43	35	40
	197	49	35	72	45	38	43	35	40
	198	50	35	35	45	38	43	35	40
	199	50	35	37	45	38	44	35	40
	200	50	36	37	45	38	93	35	40
	201	50	37	37	45	39	45	36	40
	202	50	39	74	45	38	44	35	40
	203	50	36	37	45	38	44	35	40
	205	110	39	36	45	38	44	35	40
	206	110	39	36	45	38	81	35	40
	210	96	37	37	45	39	77	36	40
	211	96	37	37	45	39	79	36	86
	212	96	75	36	98	84	44	78	86
	213	96	75	36	98	38	82	78	86
	247	50	36	82	45	38	44	35	40
	248	110	35	37	45	38	44	35	40

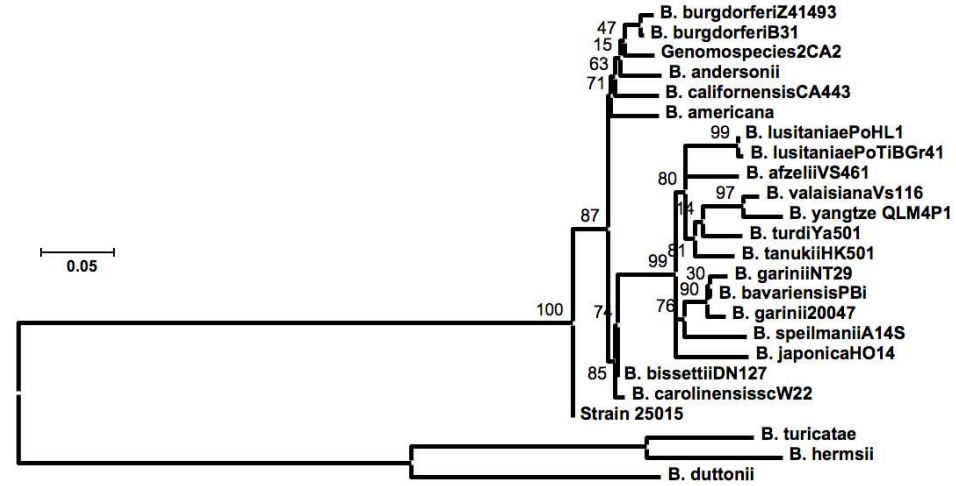
Table A4.2 Allelic profiles identified in Latvian ticks from 2002

Strain	Year	Tick Stage	Species	Site	ST	*clpA*	*clpX*	*nif*	*pep*	*pyr*	*rec*	*rpl*	*uvr*
24703LT	2002	nymph	*B. afzelii*	Jaunciems	7701	35	24	23	86	22	27	23	28
21006LT	2002	nymph	*B. afzelii*	Kemeri	204	109	24	23	89	22	27	23	28
20809LT	2002	nymph	*B. afzelii*	Babite	7705	36	82	24	88	92	27	23	28
2459LT	2002	nymph	*B. afzelii*	Babite	7703	36	24	24	31	22	92	23	28
21301L	2002	adult	*B. afzelii*	Jaunciems	170	37	24	24	31	22	92	23	28
25415L	2002	nymph	*B. afzelii*	Jaunciems	7708	38	24	25	32	90	29	24	28
22215L	2002	nymph	*B. afzelii*	Jaunciems	215	51	24	23	86	85	27	23	29
24315L	2002	nymph	*B. afzelii*	Jaunciems	215	51	24	23	86	85	27	23	29
25215L	2002	nymph	*B. afzelii*	Jaunciems	215	51	24	23	86	85	27	23	29
24109L	2002	nymph	*B. afzelii*	Babite	7702	36	24	23	49	26	7791	23	29
2103LT	2002	nymph	*B. afzelii*	Jaunciems	220	109	24	24	85	90	91	24	29
215LT	2002	adult	*B. afzelii*	Kemeri	220	109	24	24	85	90	91	24	29
2243LT	2002	nymph	*B. afzelii*	Jaunciems	220	109	24	24	85	90	91	24	29
23515L	2002	nymph	*B. afzelii*	Jaunciems	220	109	24	24	85	90	91	24	29
23212L	2002	nymph	*B. afzelii*	Babite	7709	109	24	23	31	23	30	23	30
253LT	2002	nymph	*B. afzelii*	Jaunciems	7706	37	24	23	31	22	92	23	30
25815L	2002	nymph	*B. afzelii*	Jaunciems	7707	37	24	25	31	7796	92	23	30
20211L	2002	adult	*B. afzelii*	Babite	7703	36	24	23	87	92	27	23	78
21621L	2002	nymph	*B. afzelii*	Babite	7703	36	24	23	87	92	27	23	78
2621LT	2002	nymph	*B. afzelii*	Babite	7710	8836	24	23	87	92	27	23	78
22521LT	2002	nymph	*B. burgdorferi*	Babite	20	14	1	11	1	1	1	1	10
264LT	2002	adult	*B. burgdorferi*	Kemeri	20	14	1	11	1	1	1	1	10
21509LT	2002	nymph	*B. burgdorferi*	Babite	21	14	1	11	1	1	10	1	10
20111LT	2002	adult	*B. burgdorferi*	Babite	161	15	9	12	8	1	17	8	16
25103LT	2002	nymph	*B. garinii*	Jaunciems	7713	48	82	24	88	7792	27	23	28
24912L	2002	nymph	*B. garinii*	Babite	87	42	27	29	38	29	39	27	33
22112L	2002	nymph	*B. garinii*	Babite	87	42	27	29	38	29	39	27	33
20219L	2002	adult	*B. garinii*	Babite	180	43	28	30	90	87	36	28	34
20510L	2002	adult	*B. garinii*	Babite	180	43	28	30	90	87	36	28	34
20322L	2002	adult	*B. garinii*	Babite	7715	7743	28	30	90	87	36	28	34
20519L	2002	adult	*B. garinii*	Babite	89	44	29	31	40	31	37	29	35
20110L	2002	adult	*B. garinii*	Babite	90	45	30	32	41	32	38	30	36
23512L	2002	nymph	*B. garinii*	Babite	187	47	73	33	42	91	76	32	36
24106L	2002	nymph	*B. garinii*	Kemeri	187	47	73	33	42	91	76	32	36
24909L	2002	nymph	*B. garinii*	Babite	187	47	73	33	42	91	76	32	36
2416LT	2002	adult	*B. garinii*	Kemeri	7712	47	7701	33	42	91	7776	32	36
2286LT	2002	nymph	*B. garinii*	Kemeri	244	48	34	34	44	37	42	33	39
24309L	2002	nymph	*B. garinii*	Babite	244	48	34	34	44	37	42	33	39
227LT	2002	adult	*B. garinii*	Babite	7711	44	29	76	40	31	37	80	77
20101L	2002	adult	*B. garinii*	Jaunciems	251	95	74	34	96	89	78	77	85
21721L	2002	nymph	*B. garinii*	Babite	7714	95	7702	34	96	89	78	77	7785
21617L	2002	adult	*B. lusitania*	Kemeri	7700	n101	n22	n20	n27	18	7725	n74	n27
21821L	2002	nymph	*B. valaisiana*	Babite	7730	50	35	37	45	38	44	35	40
20719L	2002	adult	*B. valaisiana*	Babite	102	50	39	36	45	38	44	35	40

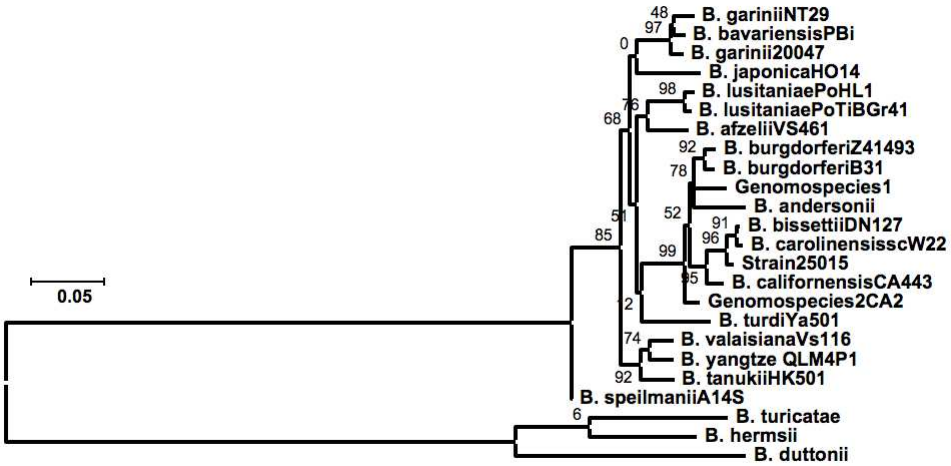
Appendix 5



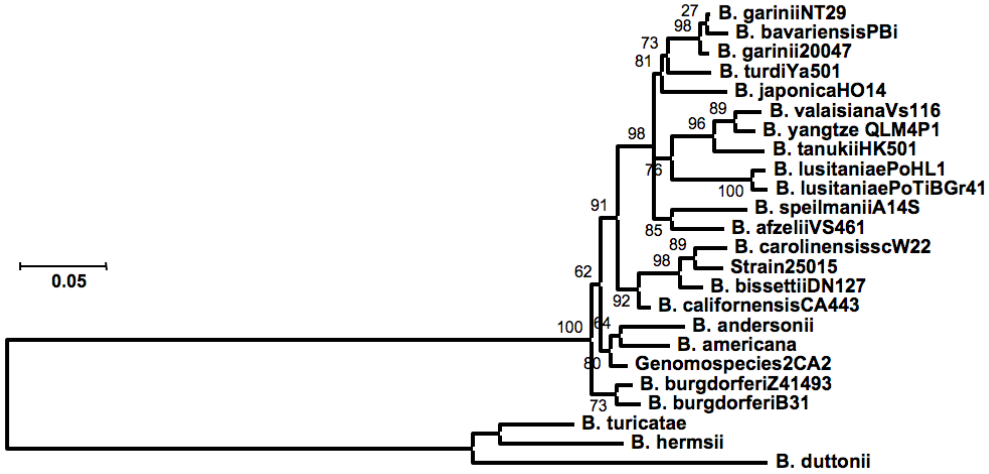
A: clpA



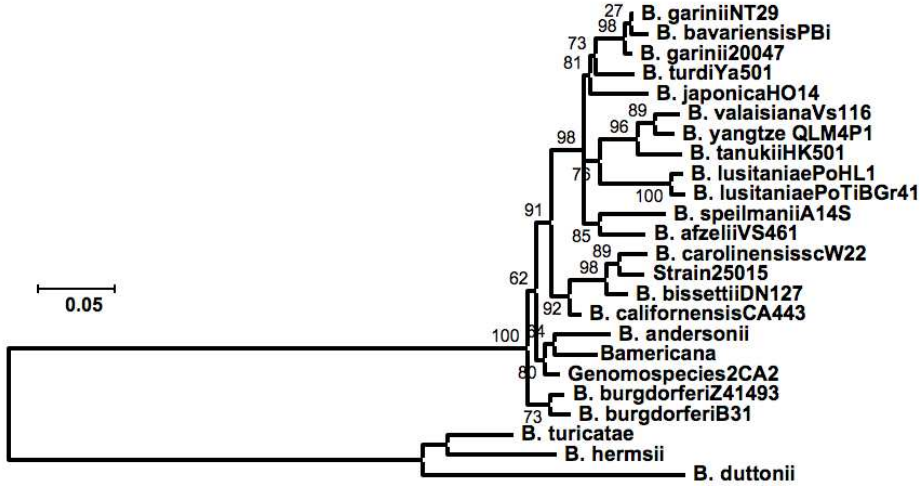
B: clpX



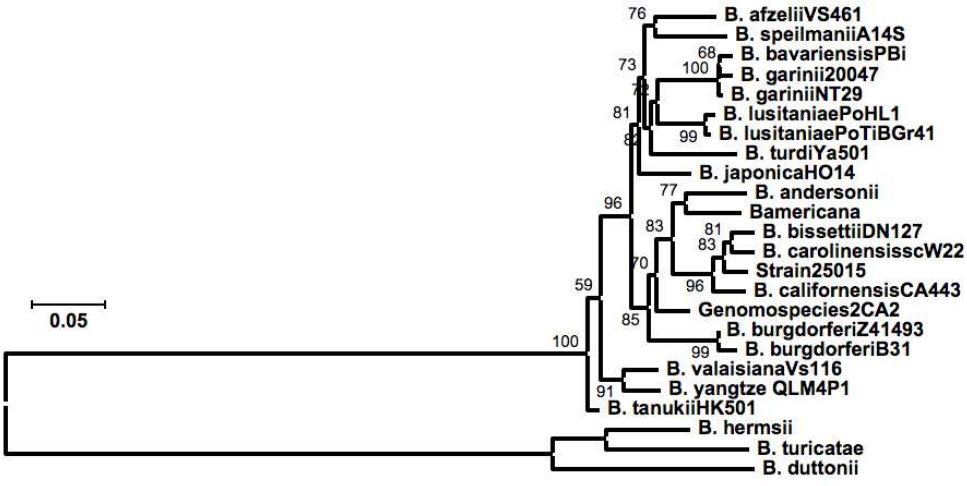
C: nifS



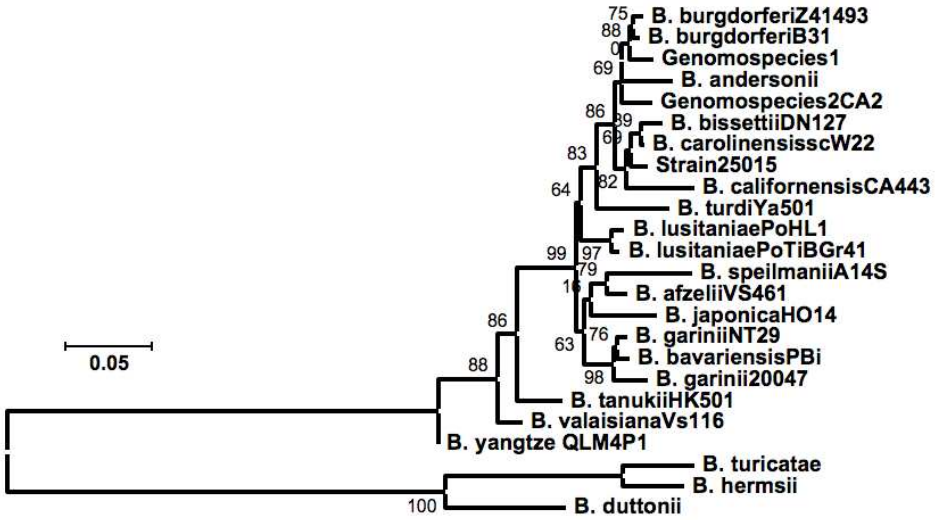
D: pepX



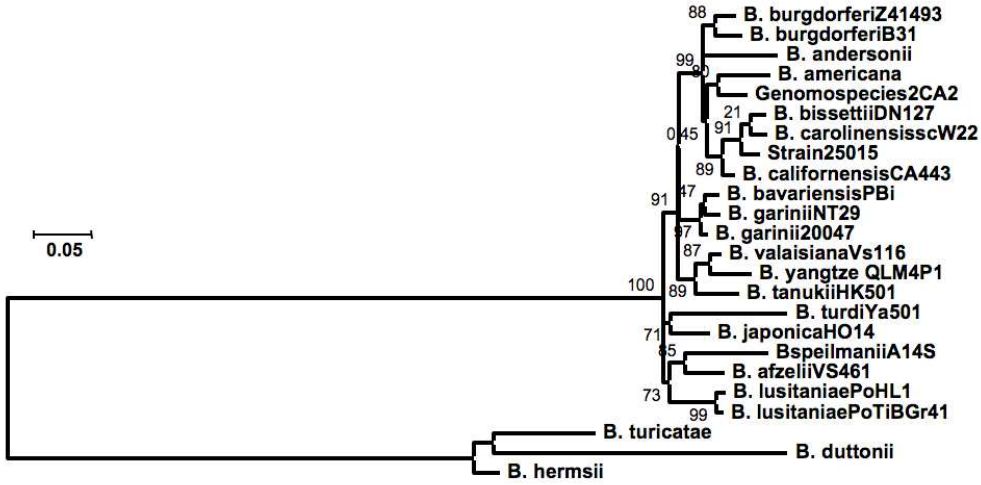
E: pyrG



F: recG

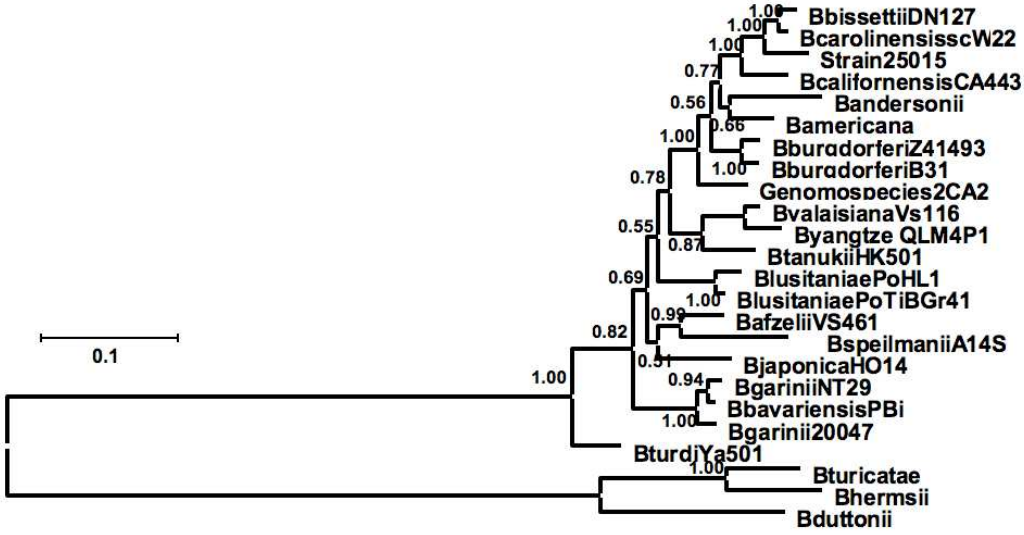


G: rplB

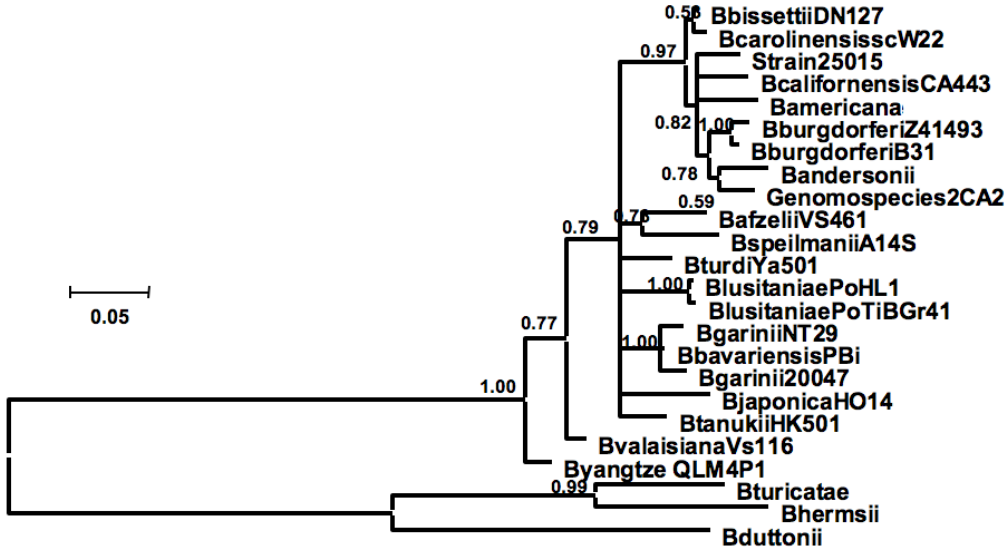


H: uvrA

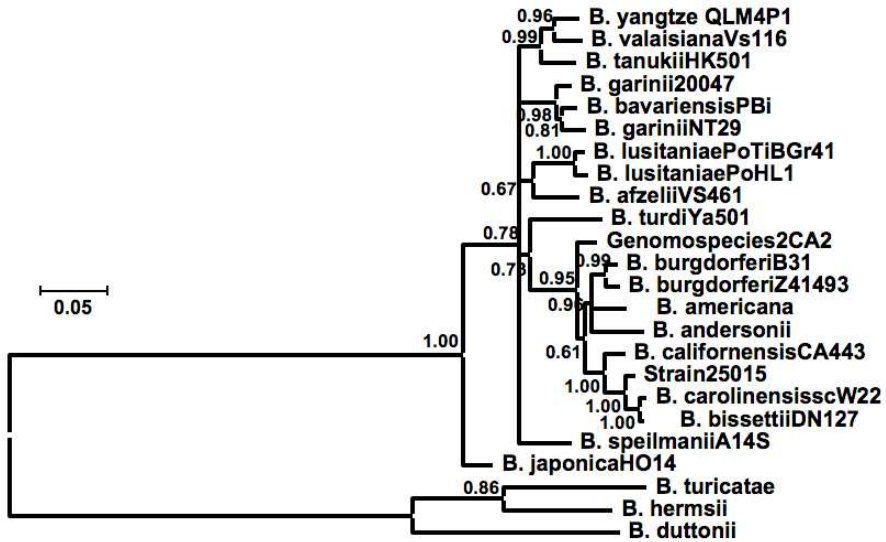
Figure A5.1 PhyML phylogenetic inferences of the individual house keeping genes (A-H) using the LB group of species. The outgroup species are the relapsing fever spirochaetes and branch support values were calculated using aLRT. Scale bars represent either 5 or 10% divergence.



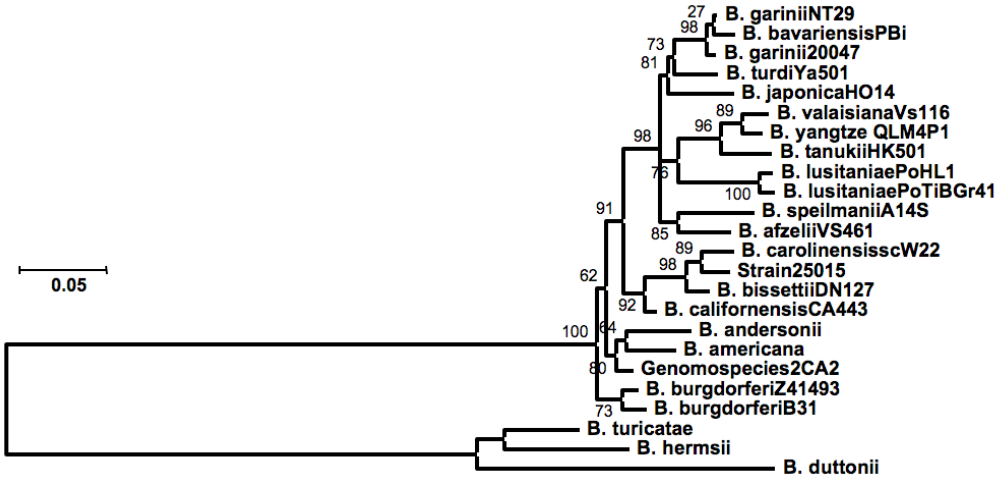
A: clpA



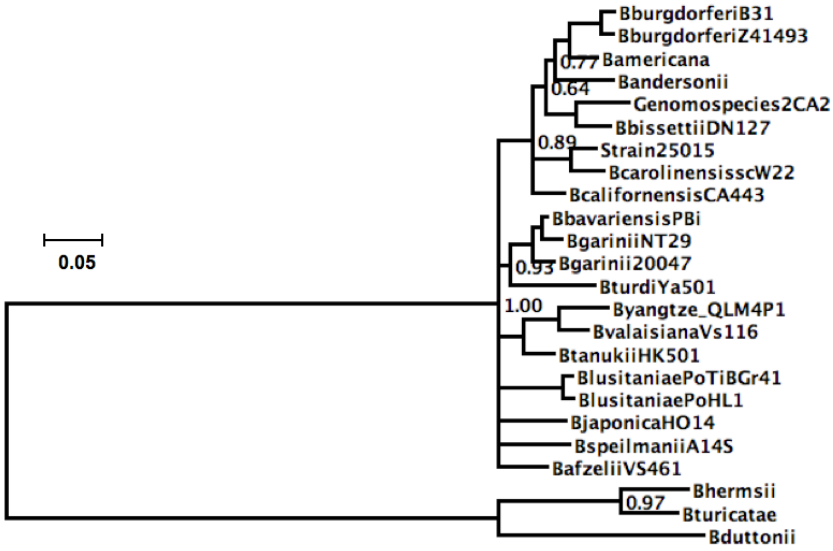
B: clpX



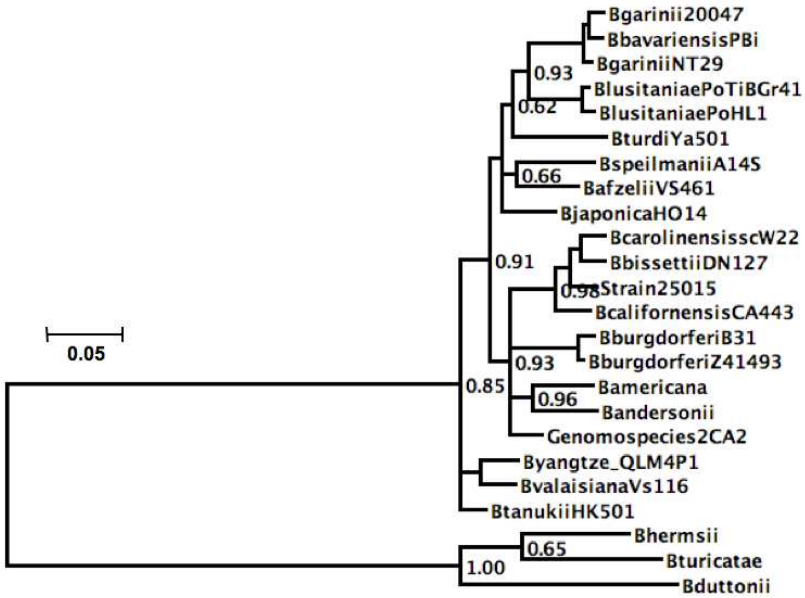
C: nifS



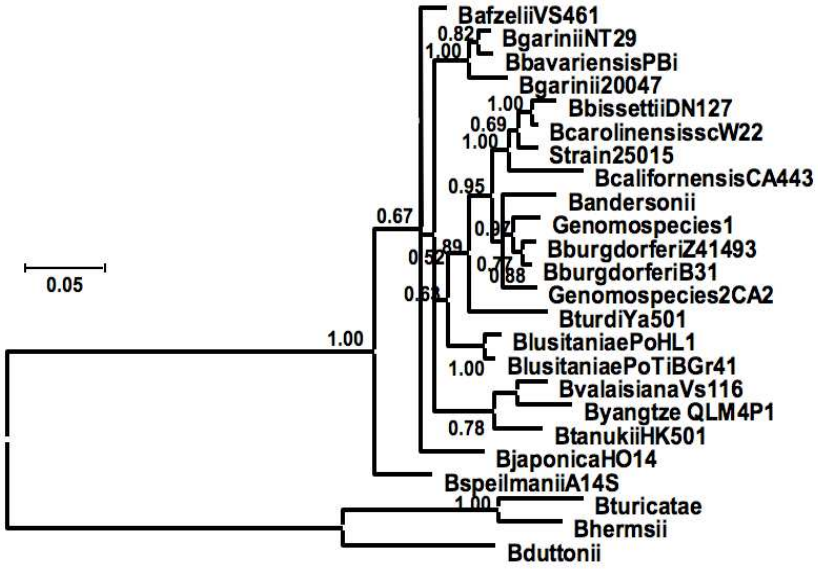
D: pepX



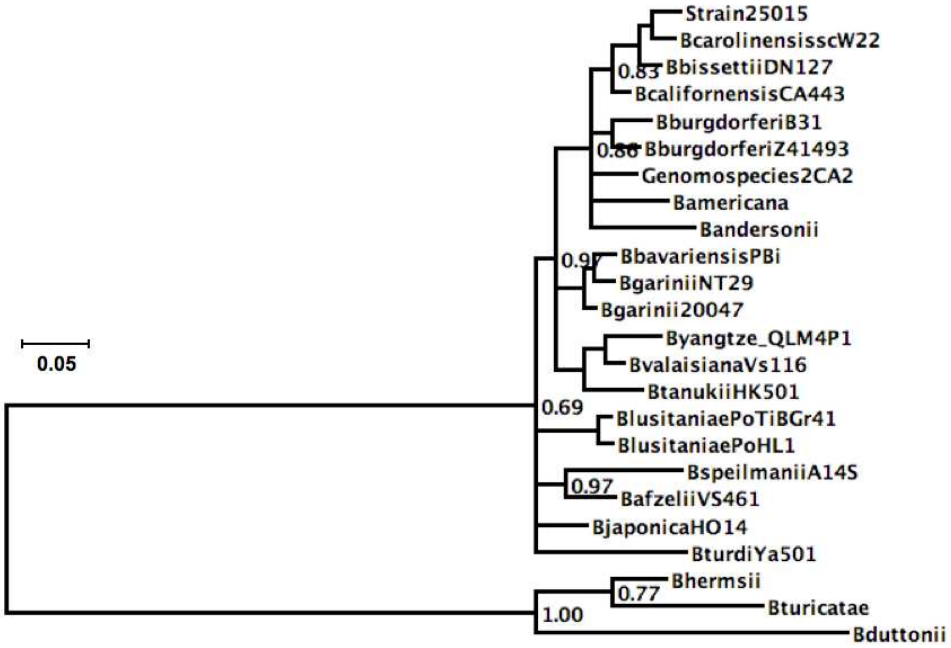
E: pyrG



F: recG



G: rplB



H: uvrA

Figure A5.2 Bayesian phylogenetic inferences of the individual house keeping genes (A-H) using the LB group of species. The outgroup species are the relapsing fever spirochaetes and branch support values were calculated using posterior. Scale bars represent either 5 or 10% divergence.

Appendix 6

Table A6.1 Output from TREE-PUZZLE showing the probability that the alignment accepts the tree topologies for *clpA, clpX, nifS, pepX, pyrG, recG, rplB, uvrA* using the PhyML inference method and the concatenated gene topologies using either the PhyML or Bayesian inference method. The four tests used to judge whether the alignment accepts the topology are the one-sided Kishino-Hasegawa test (p-1sk), the Shimodaira-Hasegawa test (p-SH), the expected likelihood weights (c-ELW) and the two-sided Kishino-Hasegawa test (2sKH). The word 'best' indicates which tree the alignment best fits.

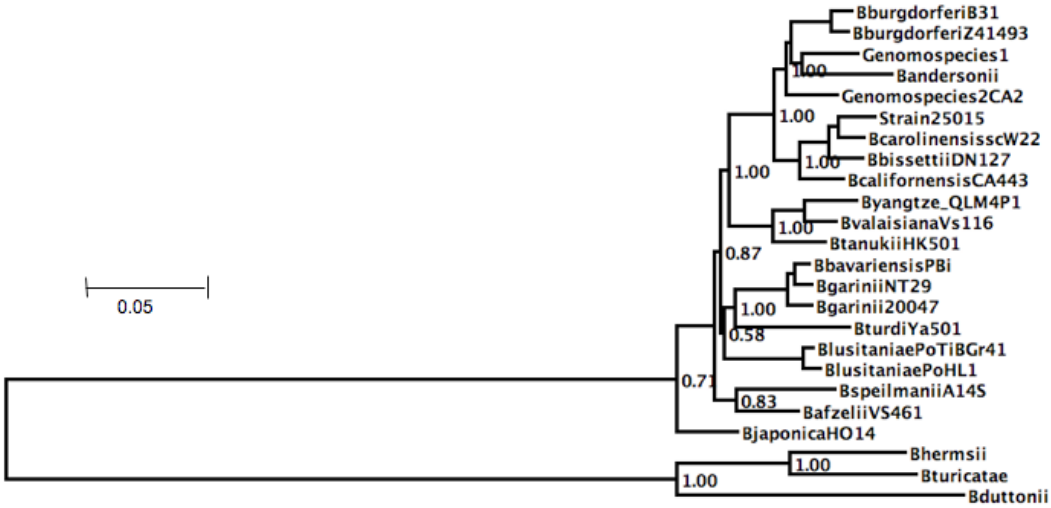
***clpA* alignment**										
Tree	**log L**	**difference**	**S.E.**	**p-1sK**		**p-SH**		**c-ELW**		**2sKH**
clpA	-5382.74	0.00 <	best	1	+	1	+	0.6287	+	best
clpX	-5495.66	112.92	31.3905	0.001	-	0.002	-	0	-	-
nifS	-5414.62	31.88	15.9315	0.021	-	0.279	+	0.0049	-	-
pepX	-5421.72	38.98	20.5473	0.026	-	0.152	+	0.0023	-	+
pyrG	-5510.57	127.83	28.2311	0	-	0	-	0	-	-
recG	-5400.91	18.17	13.678	0.082	+	0.523	+	0.0439	+	+
rplB	-5406.54	23.8	17.1323	0.087	+	0.434	+	0.0155	-	+
uvrA	-5425.91	43.17	19.3159	0.024	-	0.113	+	0.0018	-	-
concat. PhyML	-5391.25	8.51	15.2149	0.308	+	0.807	+	0.1862	+	+
concat. Bayes	-5392.26	9.52	15.2061	0.257	+	0.792	+	0.1165	+	+
***clpX* alignment**										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**
clpA	-4765.94	8.79	27.1883	0.377	+	0.784	+	0.1315	+	+
clpX	-4757.15	0.00 <	best	1	+	1	+	0.498	+	best
nifS	-4764.76	7.6	28.1148	0.378	+	0.725	+	0.2386	+	+
pepX	-4792.36	35.21	26.4289	0.09	+	0.16	+	0.002	-	+
pyrG	-4820.06	62.91	24.1731	0.006	-	0.016	-	0	-	-
recG	-4776.6	19.45	28.1599	0.23	+	0.471	+	0.0187	-	+
rplB	-4773.14	15.99	26.2662	0.265	+	0.613	+	0.0347	+	+
uvrA	-4780.64	23.48	29.1495	0.228	+	0.394	+	0.0474	+	+
concat. PhyML	-4773.81	16.66	25.6755	0.27	+	0.572	+	0.0075	-	+
concat. Bayes	-4771.99	14.84	25.9866	0.259	+	0.628	+	0.0217	-	+
***nifS* alignment**										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**
clpA	-4416.6	23.6	11.4411	0.023	-	0.317	+	0.0088	-	-
clpX	-4521.05	128.05	27.2194	0	-	0	-	0	-	-
nifS	-4392.99	0.00 <	best	1	+	1	+	0.9294	+	best
pepX	-4436.24	43.24	20.9611	0.023	-	0.067	+	0.014	-	-
pyrG	-4509.42	116.43	26.0052	0	-	0	-	0	-	-
recG	-4422.07	29.07	14.5194	0.03	-	0.213	+	0.0091	-	-
rplB	-4415.98	22.99	12.892	0.04	-	0.338	+	0.0242	+	+
uvrA	-4438.34	45.34	15.1698	0.003	-	0.047	-	0	-	-
concat. PhyML	-4422.96	29.97	14.6756	0.019	-	0.178	+	0.0006	-	-
concat. Bayes	-4419.72	26.72	15.3922	0.047	-	0.234	+	0.0139	-	+
***pepX* alignment**										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**

clpA	-4661.33	50.02	21.2711	0.019	-	0.047	-	0.0025	-	-
clpX	-4767.45	156.13	31.3344	0	-	0	-	0	-	-
nifS	-4663.88	52.56	21.3211	0.012	-	0.03	-	0	-	-
pepX	-4611.32	0.00 <	best	1	+	1	+	0.7329	+	best
pyrG	-4695.83	84.51	23.8272	0	-	0.001	-	0	-	-
recG	-4647.8	36.48	18.4116	0.033	-	0.17	+	0.0078	-	-
rplB	-4647.2	35.88	18.7539	0.025	-	0.163	+	0.0002	-	+
uvrA	-4639.32	28	18.2691	0.073	+	0.32	+	0.034	+	+
concat. PhyML	-4625.85	14.54	14.3897	0.163	+	0.651	+	0.0182	-	+
concat. Bayes	-4622.06	10.74	15.5484	0.237	+	0.75	+	0.2044	+	+
pyrG alignment										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**
clpA	-4551.6	51.46	26.3262	0.017	-	0.033	-	0.0208	-	+
clpX	-4592.96	92.81	24.6491	0	-	0	-	0	-	-
nifS	-4563.79	63.65	21.7752	0.005	-	0.001	-	0	-	-
pepX	-4537.93	37.79	17.8671	0.02	-	0.08	+	0.0067	-	-
pyrG	-4500.14	0.00 <	---- best	1	+	1	+	0.9634	+	best
recG	-4557.05	56.91	26.3354	0.012	-	0.014	-	0.0054	-	-
rplB	-4558.21	58.07	21.8352	0.006	-	0.005	-	0.0009	-	-
uvrA	-4565.11	64.97	23.2242	0.001	-	0.003	-	0	-	-
concat. PhyML	-4547.96	47.82	20.7825	0.008	-	0.023	-	0.0019	-	-
concat. Bayes	-4550.06	49.92	20.7194	0.01	-	0.014	-	0.001	-	-
recG										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**
clpA	-4380.6	15.98	13.5366	0.135	+	0.526	+	0.0916	+	+
clpX	-4481.66	117.04	26.4154	0	-	0	-	0	-	-
nifS	-4400.75	36.13	16.8451	0.012	-	0.092	+	0.0015	-	-
pepX	-4390.56	25.93	15.1464	0.044	-	0.27	+	0.0071	-	+
pyrG	-4461.6	96.98	22.4754	0	-	0	-	0	-	-
recG	-4364.63	0.00 <-	best	1	+	1	+	0.6471	+	best
rplB	-4395.33	30.71	17.8807	0.051	+	0.164	+	0.0051	-	+
uvrA	-4397.9	33.28	17.005	0.026	-	0.144	+	0.0046	-	+
concat. PhyML	-4371.53	6.9	11.6916	0.264	+	0.779	+	0.2277	+	+
concat. Bayes	-4381.53	16.9	13.9346	0.118	+	0.497	+	0.0153	-	+
***rplB* alignment**										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**
clpA	-3727.7	14.07	11.178	0.106	+	0.552	+	0.0302	-	+
clpX	-3791.12	77.49	24.1896	0	-	0.006	-	0	-	-
nifS	-3726.81	13.18	13.2897	0.149	+	0.581	+	0.1019	+	+
pepX	-3726.88	13.25	15.7287	0.186	+	0.569	+	0.1197	+	+
pyrG	-3760.68	47.05	21.0533	0.011	-	0.047	-	0.0015	-	-
recG	-3737.88	24.25	11.471	0.025	-	0.261	+	0	-	-
rplB	-3713.63	0.00 <	best	1	+	1	+	0.533	+	best
uvrA	-3736.34	22.71	10.4906	0.015	-	0.285	+	0.0001	-	-
concat. PhyML	-3720.77	7.15	11.4145	0.256	+	0.779	+	0.0705	+	+
concat. Bayes	-3720.14	6.51	11.5404	0.271	+	0.791	+	0.143	+	+
***uvrA* alignment**										

Tree	log L	difference	S.E.	p-1sKH		p-SH		c-ELW		2sKH
clpA	-5086.55	38.74	22.4564	0.045	-	0.153	+	0.0073	-	+
clpX	-5166.14	118.34	32.3723	0.001	-	0.001	-	0	-	-
nifS	-5085.01	37.21	26.7711	0.088	+	0.218	+	0.034	+	+
pepX	-5080.53	32.73	27.6008	0.125	+	0.261	+	0.0719	+	+
pyrG	-5138.59	90.79	31.0999	0.005	-	0.007	-	0	-	-
recG	-5081.32	33.52	20.6086	0.058	+	0.261	+	0.0199	-	+
rplB	-5077.42	29.62	21.1904	0.092	+	0.315	+	0.0138	-	+
uvrA	-5047.8	0.00 <	best	1	+	1	+	0.768	+	best
concat. PhyML	-5068.94	21.14	19.6002	0.132	+	0.504	+	0.0349	+	+
concat. Bayes	-5067.31	19.51	18.9333	0.152	+	0.545	+	0.0502	+	+
Concatenated gene alignment										
Tree	**log L**	**difference**	**S.E.**	**p-1sKH**		**p-SH**		**c-ELW**		**2sKH**
clpA	-33708.34	27.04	35.4154	0.201	+	0.715	+	0.2021	+	+
clpX	-34294.14	612.85	70.7485	0	-	0	-	0	-	-
nifS	-33787.24	105.95	39.3806	0.005	-	0.09	+	0	-	-
pepX	-33717.85	36.56	31.9052	0.126	+	0.607	+	0.0948	+	+
pyrG	-34118.6	437.31	59.3219	0	-	0	-	0	-	-
recG	-33749.76	68.47	31.6415	0.013	-	0.288	+	0.0016	-	-
rplB	-33741.02	59.72	25.6831	0.014	-	0.376	+	0.0035	-	-
uvrA	-33772.89	91.59	34.5707	0.001	-	0.15	+	0.0012	-	-
concat. PhyML	-33681.3	0.00 <	best	1	+	1	+	0.5615	+	best
concat. Bayes	-33686.92	5.63	5.9695	0.169	+	0.919	+	0.1354	+	+

Appendix 7

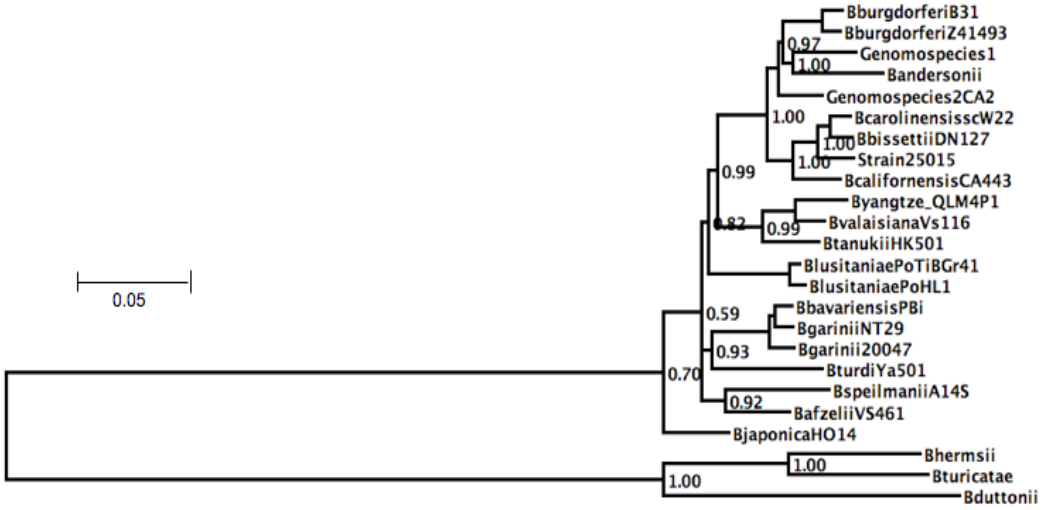
A: PhyML



A: Bayesian

Figure A7.1 Phylogenetic inferences of the MLSA housekeeping genes excluding *clpX*. Phyologeny A shows the PhyML inference and branch support values were calculated using aLRT. Phylogeny B shows the Baysian inference and branch support values were calculated using posterior probabilities. Scale bars show 5% divergence.

A: PhyML



B: Bayesian

Figure A7.2 Phylogenetic inferences of the MLSA housekeeping genes excluding *pyrG*. Phyologeny A shows the PhyML inference and branch support values were calculated using aLRT. Phylogeny B shows the Baysian inference and branch support values were calculated using posterior probabilities. Scale bars show 5% divergence.

Appendix 8

Concatenated gene sequence alignment of all LB group species

>Byangtze QLM4P1

AAAAAAGAACTTATTATAGAGGATAGTTTGGTATTTGATTTGATATTAAATATTAAATTATTAAAATTTAATTTGCTTGCCAATAG
AAGTACTATTGGTTTATTTGCCTTTATTGGTGCTTCTGGGGCTGGAAAATGCAAATTGACGGATATTTTATCGGAAGAGTTTAAAA
TTCCGAAATTTAGCCTTAATATGAGTGAATATAGCGATTTTAATGCTCTTGATAGATTGATTGGGCCTGTGTTAAGTAATGAGGGA
TATTACGAATCTACTAGATTTTTCAAATTTTTGAACAAATCTTCTAATTCTATCATTTTTTTATCAGATTTTGATAAATGTAATAAA
AGGGTTTTAGAATTTTTTTTGGAGGGGTTTAAAACAGGCAAACTTTTTGATGGTCTTGGTAAAAAGGTAAGTTTATCAGAAAGTTT
AATAGTAATAAGTGTAAATGCTGAGGGCAATGAGCTTAATAGCATCGGCTTTAAAAATAAAATAGCAGAAAGAAATGATGTTGA
TTTTATATTAAAGAAAAGGTTTCCCAATGAATTTTTACATTTAATAGATCATGTGTTTGTATTTGCTGGATATGTTGGTGAAGATGT
AGAAAATATTTTACTTAAATTAATACATGCTGCTCATGGAGATGTTAGTCTAGCTGAGAAAGGGATTATTTATATAGATGAAATAG
ATAAAATTGCTAAAAAAAAATGAAAACGTTTCAATAACAAGGGATGTTTCTGGAGAAGGGGTCCAACAGGCTTTGTTAAAGATAAT
TGAGGGTACTATTGCTAATGTTCCTCCAAGAGGAGGAAGAAAACATCCTTATGAAGACACTATTGAAATTAATACTCAAAATATA
CTCTTTATATGTGGCGGTGCATTTGTTGGGCTTGAAAATATTGTTAAGAATCGAATGAATAAAAGTTCTATTGGGTTTTCATCAACT
GAAAAAAAGAATATAAGAGAAGACACTTCATTTAAGTATTTAGAAATGGAAGATTTGATTAAATTTGGTTTGATACCAGAATTTG
TTGGCAGACTTCCTGTACATTCGTATCTTGAAAAATTAAATAAAGAAGATTTGATGAAAATATTGATTGATCCTCAAAATTCTATT
GTTAAGCAGTATTATCATATGTTTAAAATGGACAATGTTGAATTAGTATTTGAAAAAGATGCTTTAGAATCAATCGTAGATGAAGC
TTTTAACAACAAACATATAATTTATTTTGACAACGCAGCAACCTCTCAAAAACCTAAAAAAGTAATTTACTCTAGCATAGAATATT
ATGAAAATTACAACGCAAATGTACATAGAAGCGGTCACAAATTTGCAATTCAATCTAGCATAACAATAGAAAAAACAAGAGAAC
TTGTAAAAAATTTTATTAATGCAGAATCTGCAAAAAATATAATATTTAACTCTGGAACTACAGATGGAATTAACACCATCGCAAA
CTCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAGATTATTCTTACAACTCTTGAACATAATAGTAATTTGCTGCCATGGG
CAAATCTTGCAAAATTATCTAATCTAACAATTAAGTTTGCTAAATTCAATGAAATGGGAATTATTACCCCAGAAGAAATTGAAAA
ACTTATTACAGAAAAAACAAAACTCATCAGTATTTCAGGAATAAATAATACTTTGGGAACCATTAATGATTTAGAATCTATTGGG
AAAATCGCAAAAAAATATAATATAAGTCTTTTTATAGATGCTGCACAAATGGCATATAATACTCATGACAATTTAACAGTAATAA
GCAGCACTAAAAAATCAATAAAAGATAGCATACTAGAACAACTTGAAATAGAATATGAAAATTTTTTATCTTGTGATCTAATATT
CACAGAATCACAACCTTCCAAAATAATAGGAACTGAAAAAGAATTTTTGACTCTAAAAATCTTGACAATAAATCAGGATGCCAT
GCAATCATGAACTCTTATATTCATACAAGCAATAATAAAAATAAAATGGCTGTATTTTTTGATAACGAAGAAGTAGGATCTCTAA
CCTCAAGAGGCGCTGATTCCAATTTTTTATCAGAGGTTTTAGAAAGAATTGATCTTGCTCTTAACTTAACCAGAGAAGAACATTTA
ATAAAAACAAACAAATCATTTAATATATCGTTTGACAGTGTTCACGGAATTCATCCAGGATATGCATGTAAACATGATCCAAACT
ATCAAGCAACTCTAAGTAAAGGTATAGTTGTAAAAAACAGCGCCAATTTCAGATATGCAACAACTTCAACAGGATTTGCAAAATT
AAAAAATTTGGCTGTTAAAAATAATATTGGAAGTAGCAATATTTCTTTTATTCATTTAACCTATGTTCCAAGTCCGACTGGAATTA
ATGAGCAAAAATCTAAACCCACTCAACAGAGTGTTAAAACTTTAAATAAGGCAGGCATTTTCCCTGATTTGATTATTGCGAGAAG
CTCTCAAGTATTAACAGACCAAATCAGGAAAAAGGTGGCAATGTTTTGTAATGTCAAGAGCACTTCTATTATTGATAATATTGATG
TTTCTACTATTTATGAAATTCCTATATCTTTTTATAAGCAAGGTGTACATGAAATTTTAAGTTCTAAGTTAAATATTAAGGTT---
GATCCGAAAATAGAAGAGCTTTCAAAGCTTGTAGAAGTTATAAAATCTAATTTTTTTATGCCTAAAAAAATTATTAATATTGCTAT
TTGCGGTAAATATGCCGAACTTGATGATTCTTATGCGTCAATTAGAGAGTCTTTAGTCCATGTTGCAGCCAATTTGGATTTACTTAT
TAAAAGCACACTAATTGATTCTAATGATTTAAATGAGAGCTGTTTAAGAGGTTTTGATGGCATTATTGTTCCTGGTGGTTTTGGAG
GCAAAGGGTATGAGGGTAAAATCATTGCTATTGCTCGCCAACATTATGATAATTTATCCTACATATTGTCGTCTTTTGATGTTTCAG
TGACTCTTTTGACTGGCAGTTTAAAAAAGAAGGACAAAGAGCAAGCTTTAGAAAGCATTAAAAATGGAACTTCTGGTTTAATAAT
CGGAACACATGCTATTTTTTACGAAAGTACAGAATTTAAAAAATTGGCATATGTTATCATTGACGAGCAGCATAAATTTGGAGTTG
TTCAAAGAGAAGAGCTTAAAAATAAGGGAGAAGGGGTGGACATGCTTTTAATGTCTGCAACGCCTATTCCTAGAAGCTTTGCGTT
AACACTTTTTGGTGATCTTGAAATTTCGTTTATTAAAACTTTACCTAAGGGGCGTTTGCCCATTACTACTTATTTAGCAAAGCATGG
CAATGAAGATAAAGTTTATGAGTTTTTAAAAAAAGAGCTCGTAAAGGGTCATCAAGTTTATTTTGTTTATCCATTAATTTCATCTTC
AGAAAAATTTGAATTAAAAGACGTTAATAATATGTGTTTAAAATTGAAGGAAGTGTTTAGCGAATATGTTGTTGACATGCTTCATT
CTAAGTTACCATCTGATTTAAAAGAAGAAATTATGAAAAATTTTTATTCGAAAAAAGTAGATATTTTGGTGGCTACTAGTGGTAAT
GATCCTTTAAAATCTTTAACAAAAGGTAAAAAATTTAAATCAGGCAGAGATTCTTCTGGTAGGATTACTGTTAGAAGAAGGGGTG
GTGGACATAAGAGAAAGTATAGGTTGATTGATTTTAATAGAAGAGATAGATTTAGTATTCCTGCTCGAGTTGCTTCTATTGAATAT
GATCCAAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGACGGAGAAAAAAGGTATATTATTTCTCCAAAAGGCATTAAGG
TTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATAAAAATTGGTAATGCTTTGCCCCTTGAAAATATTCCTATTGGGAGAACC
ATCCACAATATTGAGCTTAATATAGGAAAAGGTGGGCAACTTGTAAGAAGTGCTGGTGGGTATGCTATGATACTTGCTTCTGACG
GAAATTATGTTACTGTAAAATTATCATCTGGTGAGATGAGGTTAATTTTTAAAAAATGTATTGCAACTATTGGTGAGGTTGGAAAT
GAAGACTACGTCAATGTTTCTATAGGGAAAGCTGGTAAAAGTAGATGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGACCATCCGCATTCGGGTAGTCTTTCTGGAGGTGAGGCTCAGCGTATTAGGCTTGCTACCCAAATAGGATCAGCACTC
TCAGGTGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAATGAAAAATTAATCTCTACTCTTGTTAACCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAACATGATGAACAAACTTTGCGTACTGCGGATTATATTATTGATATGGGGCCTGG
TGCTGGAATTCTTGGAGGGGAAATAGTTGCAAAAGGAACTTTAATTGATATTTTAAATAGCAAAAATAGTTTAACTGGGCAATAC
TTAAGTGGCAAGTTTAAAATAGATGTTCCAAGATCTAGAAGAAAGGCAGATAAGGGAGAAATTTTGCTTTTAGGTTCTAATAAAA
ACAATCTTAAGAATATAGACGTAAGTATTCCTTTGGGAGTTTTTACTGTAATAACAGGTGTTTCTGGTAGTGGAAAAAGTACTTTA
CTTAACGAAGTATTATATCCAGCTCTTGATAGCAGATTAAAGCTTAATGAAAAGTATTGTGATGGTTTTAAAGAG

>BtanukiiHK501

AAAAAAGAGCTTATTGTAGATGATAGCTTGGTATTTGATTTGATATTAAATATTAAATTATTAAAATTTAATTTGCTTGCCAATAG
AAGTACTATTGGCATATTTGCCTTTATTGGCGCTTTTGGAGTCGGAAAATGCAAATTGACGAATATTTTATCAGAAGAGTTTAAAA
TTCCAAAATTTAGTCTTAATATGGGCGAATATGGAGATTTTAATGCTCTTGATAGATTGATTGGCCCCGTTTTAAGTAATGAGGGG
TATTATGAATCTACTGGATTTTTCAAATTTTTAAACAAATCTTCTAATTCTATTATTTTTCTATCAGATTTTGATAAATGTAATAAAA
GGGTTTTAGATTTTTTTTTGGAGGGGTTTAAAACAGGTATACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCAGAAAGTTTA
ATAGTAATAAGTGTAAATGCTGAGAGCAATGAGCTTAATAGCATCGGCTTTAAAAATAAAATGGCAGAGGGAAATGATTTTAATT
TTATGCTGAAGAAGAGATTTCCCAATGAACTTTTGGAGTTAATAGATCATGTGTTTGTATTTGCTGGATATGTTGGTGAAGATGTA
GAAAATATTTTGCTTAAATTAATACATGCTGCTCATGGAGATGTTAGTTTAGCCGAGAAAGGGATAATTTATATAGATGAAATAG
ATAAAATTGCTAAAAAAAAATGAAAATGTTTCAATAACAAGAGATGTTTCTGGAGAAGGAGTTCAACAGGCTTTGTTAAAGATAAT
TGAGGGTACTATTGCTAATGTTCCCCCAAGAGGTGGTAGAAAACATCCTTATGAGGATACTATTGAAATTAATACTCAAAACATA
CTCTTTATATGTGGTGGTGCTTTTGTTGGACTTGAAAATATTGTTAAGAATCGAATAAATAAGAGTTCTATTGGGTTTTCAGCAATT
GAAAAAAAGAATATAAGAGAGGATACTTCATTAAAGTATTTAGAAATGGAAGATTTGATTAAATTTGGTTTGATACCAGAATTTG
TTGGTAGACTTCCTGTACATTCGTATCTTGAAAAATTAAATAAAGAAGATTTGATGAGAATATTAGTTGATCCCCAAAATTCTATT
GTTAAGCAGTATTATCATATGTTTAAAATGGATAATGTTGAATTAGTATTTGAAAAAGATGCTTTAGAATCAATTGTAGATGAAGC
TTTTAACAATAAGCATATAGTTTATTTTGACAATGCAGCAACCTCTCAAAAACCTAAAAAAGTAATTTATTCTAGCATAGAATATT
ATGAAAACTACAACGCAAATGTACATAGAAGCGGTCACAAATTTGCAATTCAATCTAGCATAACAATAGAAAAAACAAGAGAAC
TTGTAAAAAATTTCATTAATGCAGAATCTGCAAAAAATATAATATTTAACTCTGGCACTACAGATGGAATTAACACCATCGCAAA
CTCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAGCATAATAGTAATTTACTGCCATGGG
CAAATCTTGCAAAATTAGCTAATCTAACAATTAAATTTGCTAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAA
ACTTATTACGGAAAAAACAAAACTCATCAGTATTTCAGGAATAAATAATACTTTAGGAACCATTAACGATTTGGAATCTATTGGG
AAAATAGCAAAAAAATACAATATAAGTCTTTTTGTAGATGCTGCACAAATGGCATATAATACTCACGACAATTTAACAGTAATAA
GTAGCACTAAAAAATCAATAAAAGATAGCATCTTAGAACAACTTGGAATAGAATATGAAAATTTTCTATCTTGTGATTTAATATTT
ACAGAATCACAACCTTCCAAAATAATAGGAACTGAAAAAGAATTTTTATCCTCGAAAAATCTTGACAACAAATCAGGATGCCATG
CAATCATGAACTCTTATATTCATGCAAGCAATAATAAAAACAAAATGGCTGTATTTTTTGATAACGAAGAAGTAGGATCTCTAAC

CTCAAGAGGCGCTGATTCCAATTTTTTATCAGAAGTTTTAGAAAGAATTGATCTTGCTCTTAACTTAACCAGAGAAGAGCATTTAA
TAAAAACAAACAAATCATTTAATATTTCGTTTGACAGCGTTCACGGTATTCATCCAGGATACACATTTAAACATGATCCAAAATAT
CAAGCAACTCTAAGTAAAGGCATAGTTGTAAAAAACAGCGCCAATTTCAGATATGCAACAACTTCAACAGGATTTGCAAAATTAA
AAAACTTGGCTATTAAAAATAATATTGGAAGTGGCAATATTTCTTTTATTCATTTAACCTATGTGCCTAGTCCAACTGGAATTAAT
GAACAAAAATCTAAACCTACTCAACAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTCCCCGATTTGATTATTGCTAGAAGCTC
TCAAGTATTAACAGATCAAATCAGAAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGCACCTCTATTATTGATAATGTTGATGTTT
CTACTATTTATGAAATTCCTATATCTTTTTATAAACAAGGTGTACATAAGATTTTAAGTTCTAAGTTGAATATTAAAGTT---
GATCCGAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCTAATTTTTTTGCGCCTAAAAAAATTATTAATATTGCTAT
TTGTGGCAAGTATGCTGAACTTGATGATTCTTATGCATCAATTAGAGAGTCTTTGGTCCATGTTGCAGCTAATTTGGATTTACTTAT
TAAAAGCACTCTAATTGATTCTAATGATTTAAATGAAAGCTGTTTAAAAGGGTTTGACGGCATTATTGTTCCCGGTGGCTTTGGAG
GCAAAGGGTACGAGGGTAAAATTATTGCTATTGCTCGTCAACATTATGATAATTTATCCAACATATTGTCGTCTTTTAATGTTTCA
GTGACTCTTTTGACTGGTAGTTTAAAAAAGAAGGATAAGGAGCAAGCGTTAGAAAGCATTAAAAGTGGAACTTCTGGTTTAATAA
TAGGAACACATGCTATTTTTTACGAAAGTACAGAATTTAAAAGATTGGCATATGTTATCATTGACGAACAGCATAAATTTGGAGTT
GTCCAAAGAGAAGAACTTAAAAATAAAGGAGAAGGAGTGGACATGCTTTTAATGTCTGCAACACCTATTCCTAGAAGCTTTGCGT
TAACACTTTTTGGTGATCTTGAAGTTTCGCTTATTAAAACCTTGCCCAAGGGTCGTTTGCCTATTACTACTTATTTAGCAAAACATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAAAAAAGAGCTTATAAAGGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTT
CAGAAAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTGAAGGAAGTGTTTAGCGAATATGTTGTCGACATGCTTCAT
TCTAAGTTGCCATCTGATTTGAAAGAAGAAATTATGAAAAATTTTTATTCGAAAAAAGTAGATATTTTGGTTGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCAGGCAGAGATTCTTCTGGTAGGATTACTGTTAGGAGAAGAGGT
GGTGGGCATAAGAGAAAGTATAGGTTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTGCTTCTATTGAATA
TGATCCAAATAGAAGCGCTAATATAGCCTTGCTTGTTTATAAAGACGGAGAAAAAAGATATATTATTTCTCCAAAAGGCATTAAG
GTTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGTAATGCTTTGCCTCTTGAAAACATTCCTATTGGAAGAAC
TATTCACAATATTGAGCTTAACGTAGGAAAGGGTGGGCAACTTGTAAGAAGTGCTGGCGGGTATGCTATGATACTTGCTTCTGAC
GGGAATTATGTCACTGTAAAATTATCGTCTGGTGAAATGAGGTTAATTTTCAAAAAATGTATTGCAACAATTGGTGAGATTGGAA
ATGAAGATTATGTTAATGTTTCTATAGGGAAAGCTGGTAAAAGTAGATGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCTAT
GAATCCCGTTGACCATCCTCATTCGGGCAGCTTGTCTGGGGGTGAAGCTCAGCGTATTAGGCTTGCTACTCAAATAGGATCAGCAC
TTTCGGGCGTTATTTATGTTCTTGATGAGCCGAGTATTGGTCTTCATCAAAGAGATAACGAAAAATTAATCTCTACTCTTGTTAATC
TTAAAAATCTTGGCAATACTGTAATTGTTGTTGAGCATGATGAACAAACTTTGCGCACCGCAGACTATATTATTGATATGGGTCCT
GGTGCGGGAATTCTTGGAGGGGAAATAGTTGCAAAAGGAACTTTAATTGATATTTTAAATAGCAAAAATAGTTTAACTGGTCAAT
ATTTGAGTGGAAAGTTTAAAATAGATGTTCCAAGCTCTAGAAGAAAGACAGATAAGGGAGAAATTTTGCTTTTAGGGTCTAATAA
AAACAATCTTAAGAATATAGACGTAAGTATTCCCTTGGGAGTTTTTACTGTAATAACAGGCGTTTCTGGCAGTGGAAAAAGCACTT
TACTTAATGAAGTGTTATATCCGGCTCTTGATAGCAGATTAAAGCTTAATGAAAAGTATTGTGATGGTTTTAAAGAC

>BjaponicaHO14
AAAAAAGAGCTTATTATACAAGATGGTTTGGTATTTGATTTGATATTAAATATCAAACTATTAAAATTTAATATGCTTGCCAACAG
AAGTACTATTGGCGTATTTGTCTTTATTGGTGCTTCTGGAGCAGGAAAATGCAAATTGACAGATATTTATCAGAAGAGTTAAAA
TTCCGAAATTTAGTCTTAATATGAGCGAGTATAGCGATTTTAATTCTCTTGATAGATTGATTGGTCCTGTTTTAAGTAATGAAGGGT
ATTATGAATCTACTAGATTTTTCAAATTTTTGAACAAATCTTCTAATTCTATTATTTTCTTATCAGATTTTGATAAATGTAATAAAA
GGGTTTTAGATTTTTTTTTGGAGGGGTTTAAAACGGGCAGACTTTTTGATGGTCTTGGGAAAAAGGTAAGCTTATCAGAAAGTTTA
ATAATAATAAGTATCAATGCTGAGAGTAATGAGCTTAATGGTATTGGCTTTAAAAATAAAATGGCAGGAGAAAATCATTTTGATC
TTACATTGAAGAAGAGATTTCCTAATGAATTTTTAGAGTTAATAGATTATGTTTTTGTGTTTGCTGGATATGTTGGTGAAGATGTAG
AAAATATTTTGCTTAAATTGATACATGCTGCTCATGGAGATGTTAGTTTAGCTGAGAAAGGGATCATTTATATAGATGAAATAGAT
AAAATTGCTAAAAAAAAATGAAAATGTTTCAATAACAAGGGATGTTTCTGGAGAAGGAGTTCAACAGGCTTTATTAAAGATAATTG
AGGGTACTATTGCCAATGTTCCTCCAAGGGGTGGGAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATACT
CTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAGAATATTGTTAAGAATCGAATAAATAAGAGTTCTATTGGGTTTTCAGCAATTG
AAAAAAAGAATATAAAAGAAGACGCTTCATTAAAGTATTTGGAAATGGAAGATTTGATTAAATTTGGCTTGATTCCAGAGTTTGT
TGGTAGACTTCCTGTTCATTCGTATCTTGCAAAGTTAAATAAAGAAGATTTGATGAGAAATATTAGTTGATCCTCAAAATTCTATTG
TTAAGCAGTATTATCATATGTTCAGAATGGATAATGTTGAGTTGGTATTTGAAAAAGATGCTTTAGAGTCAATTGTAGATGAAGCT
ATTGACAATAAACATATAATTTATTTTGATAATGCAGCAACCTCTCAAAAGCCTAAAAAAGTAATTTACTCTAGCATCGAATATTA
TGAAAATTACAATGCAAATGTACACAGAAGCGGTCACAAATTTGCAATTCAATCAAGCATAAAAATAGAAAAAACAAGAGAACT
TGTAAAAAATTTCATTAATGCAGAATCTGCAAACAATATAATATTTAACTCTGGAACTACGGACGGAATCAATACCATTGCAAAC
TCGTTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACACAATAGTAATTTACTGCCATGGGC
AAATCTTGCAAAATTAGCTAATCTAACAATTAAGCTTGCTAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAA
CTTATTACAGAAAAAACAAAGCTTATCAGTATTTCCGGAATAAATAATACCCTAGGAACCATTAATGATTTAGAATCTATTGGAA
AAATCGCAAAAAAATATAATATAAGTCTTTTTATAGATGCTGCACAAATGGCATATAATACCCATGACAATTTAACAGTAATTAA
CAGCACCAAAAAAACAATAAAAGATAATATTTTAGAACAACTTGGAATAGAATATGAAAATTTTCTATCTTGCGATCTAATATTC
ACAGAATCACAACCTTCTAAGATAATAGGAACTGAAGGAGAATTTTTGGCTTCCAAAAATCTTGATAACAAATCGGGATGCCATG
CAATCATGAACTCTTATATTCATACAAGTAATAATAACAATAAAATAGCTGTATTTTTTGTAACGAAGAAGTAGGATCTTTAACA
TCAAGGGGCGCTGATTCCAATTTTTTATCAGAAGTTTTAGAAAGAATCGATATTGCTCTTAACTTAACCAGAGAAGAGCATTTGAT
AAAAACAAACAAATCATTTAATATATCGATTGACAGCGTTCACGGCATTCATCCAGGATATACATTTAAACATGATCCAAACTAT
CAAGCAACTCTAAGTAGAGGCATAGTTGTAAAAAATAGTGCCAATTTCAAATACGCAACAACCTCAACAGGATTTGCAAAATTAA
AAAATTTGGCTATTAAAAATAATATTGGAAGCGACAATATTGCTTTTATTCATTTAACCTATGTGCCGAGTCCAGCTGGAATTAAT
GAGCAAAAATCTAAACCCACTCAGCAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTTCCTGATTTGATTATTGCCAGAAGCTC
TCAAGTGTTGACGGATCAAATCAGAAAAAAAGTGGCAATGTTTTGTAATGTCGAGAGCACTTCTATTATTGATAATGTTGATGTTT
CTACTATTTATGAAATTCCAATATCTTTTTACAAACAAGGTGTACATGAGATTTTAAGTTCTAGGTTGAATATTAAGGTT---
GACCCAAAAATAGAGGAGCTTTCAAAGCTTGTAGGCGTTATAAAATCTAATTTTTTTGTGCCTAAAAAAATTATTAATATTGGTAT
TTGTGGCAAATATTCTGAACTTGATGATTCTTATGCTTCAATTAGAGAGTCTTTGGTTCATGTTGCAGCTAATTTGGATTTACTTAT
TAAAAGCACCTTAATTGATTCTAATGATTTAAATGAAAGCTGTTTAAAAGATTTTGACGGTATTATTGTTCCCGGTGGTTTTGGAG
GTAAAGGATATGAAGGTAAAATTATTGCTATTGCTCGCCAGCATTATGATAATTTATCCAACATATTGTCATCTTTTAACATTTCA
GTGACTCTTTTGACTGGTAGTTTGAAAAAGAGAGATAAGGATCAAGCGTTGGAAAGCATTAAAAGCGGAGCTTCAGGTTTAATAG
TTGGAACACATGCTATTTTTTACGAAAGTACAGAATTTAAAAGATTAGCATATGTTATCATTGACGAACAGCATAAATTTGGAGTT
ATTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGGGTAGACATGCTTTTAATGTCTGCAACACCTATTCCTAGGAGCTTTGCGC
TAACACTTTTTGGTGATCTTGAAGTTTCGCTTATTAAAACCTTGCCTAAGGGCCGTATACCTGTTACTACTTATTTAGCAAAGCATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAAGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTT
CAGAGAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTGAAGGAAGTGTTTAGTGAATATGTTGTCGAGATGCTTCAT
TCTAAGTTGCCATCTGATTTAAAAGAAGAAATTATGAAAAATTTTTATTCTAAAAAAGTAGATATTTTGGTGGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCGGGCAGAGATTCTTCTGGTAGGATTAGTATTAGAAGAAGAGGT
GGTGGACATAAGAGAAGATATAGGTTAATTGATTTTAATCGAAGAGATAAATTTAGCATTCCCGCTCGAGTTGCTTCTATTGAATA
CGATCCTAATAGAAGTGCTAACATAGCGTTGCTTGTTTATAAAGATGGAGAAAAAAGATATATCATTTCTCCTAAAGGCATTAAG
GTTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGCAATGCCTTGCCCCTTGAAAATATTCCTATTGGAAGAAC
TGTTCACAATATTGAACTTAATGTAGGAAAGGGTGGACAGCTTGTAAGAAGTGCTGGGGGATATGCTATGATACTTGCTTCTGAT
GGAAATTATGTCACTGTAAAATTGTCGTCTGGTGAGATGAGATTAATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAA
ATGAAGATTATGTAAATGTTTCTATAGGTAAAGCTGGCAAAAGTAGGTGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCTAT
GAATCCTGTTGACCATCCGCATTCGGGCAGTTTATCTGGTGGCGAGGCTCAGCGCATTAGGCTTGCTACTCAAATAGGGTCAGCAC
TTTCGGGCGTTATTTATGTTCTTGATGAGCCAAGTATTGGCCTTCATCAAAGAGATAACGAAAAATTAATTTCCACTCTTGTTAATC
TTAAAAATCTTGGCAATACGGTAATTGTTGTTGAGCATGATGAGCAAACTTTGCGTACTGCAGACTATATTATTGATATGGGCCCT

GGTGCTGGAATTCTTGGAGGGGAAATAGTTGCAAAGGGGGCCTTAATTGATATTTTAAATAATAAAAATAGCTTAACTGGTCAAT
ATTTGAGTGGCAAGTTTAAAATAGATGTTCCAAGTTCTAGAAGAAAGACAGATAAGGGAGAAATTTTGCTTTTAGGTTCTAATAA
AAACAATCTTAAAAATATAGACGTAAGCATTCCTTTAGGAGTTTTTACTGTAATAACAGGTGTTTCTGGTAGTGGAAAAAGCACTT
TACTTAACGAAGTGTTGTATCCGGCTCTTGATAGCAGATTAAAGTTTAATGAAAAGTATTGCGATGGTTTTAAAGAC

>BgariniiNT29
AAAAAAGAACTTATTATACACGATAGTTTAGTATTTGATTTGATACTAAATATAAAACTATTAAAATTTAATTTACTTGCCAACAG
AAGTACTATTGGTATATTTGCCTTTATTGGTCCTTCTGGAGCGGGAAAATGCAAATTGACGGATATTTTATCAGAAGAGCTTGGAA
TTCAAAAATTTAGTCTTAATATGGGCGAGTATGGCGATTTTAATTCTCTTGATAGGTTGATTGGACCTGTTTTAAGCAATGAAGGG
TATTATGAATCTACTAGATTTTTTAATTTTTTGAACAAATCTTCTAATTCTATTATTTTTCTATCAGATTTTGATAAATGTAATAAAA
GGGTTTTAGATTTTTTTTTGGAAGGATTTAAAACAGGCAAGCTTTTTGATGGTCTTGGGAAAAAGGCAAGTTTATCAGAAAGTTTA
ATAGTAATAAGTATTAATACTGAGAGTAATGAGCTTAATAGTATTGGCTTTAAAAATAAAATGACAGGGGAAAATGATTTTAATT
TTATATTAGAGAAGAGATTTCCTAATGAATTTTTAGAGTTAATAGATTATGTTCTTTTATTTGCTGGATATGTTGGTGAAGATGTAG
AAAATATTTTGCTTAAATTAATACATGCTGCTCATGGGGATGTTAGTCTGGCTGAGAAAGGGATTATTTATATAGACGAAATAGAT
AAAATTGCTAAAAAAAAATGAAAATGTTTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAGCAGGCCTTGTTAAAGATAATTG
AGGGTACTATTGCCAATGTTCCCCCAAGAGGTGGTAGAAAGCATCCTTATGAGGACACTATTGAAATTAATACTCAAAATATACT
CTTTATATGTGGTGGCGCTTTTGTTGGGCTTGAAAATATTGTTAAGAATCGAATAAACAAGAGTTCTATTGGGTTTTCGGCAATTG
AAAAAAAGAATATAAGAGAGGACACTTCATTAAAGTATTTAGAAATGGAAGATTTGATTAAATTTGGCTTGATACCAGAGTTTGT
TGGTAGACTTCCTGTACATTCGTATCTTGAAAAGTTAAATAAAGAAGATTTGATGAGAATATTAGTTGATCCTCAAAATTCAATTA
TTAAGCAGTATTATCATATGTTTAAAATGGATAATGTTGAATTGTTATTTGAAAAAGATGCTTTGGAATCAATTGTAGATGAAGCT
TTTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACCTCTCAAAAGCCTAAAAAAGTAATTTATTCTATCATTGAATATTA
TGAAAATTACAACGCAAATGTACATAGAAGCGGTCATAAATTTGCAATTCAATCTAGCATAAAAATTGAAAAAACAAGGGAACTT
GTAAAAAATTTCATTAATGCAGAATCTTCAAAAAATATAATATTTACCGCTGGAACTACGGATGGAATTAACACCATCGCAAGCT
CATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAATAGTAATTTGCTGCCATGGGCA
AATCTTGCAAAATTAGCCAACCTAACAATTAAATTTGCTAAATTCAATGAAATGGGAATTATTACCCCCGAAGAAATTGAAAAAC
TTATTACAGAAAAAACAAAGCTCATAAGCATTTCAGGAATAAATAATACCTTGGGAACAATGAATGATTTAGAATCTATTGGAAA
AATCGCAAAAAAATATAATATAAGTCTTTTTATAGATGCTGCACAAATGGCGTACAATACTCACGACAATTTAACAGTAATAAAT
AGCACTAAAAAAACAATAAAAGATAATATCTTAGAACAACTTGGAATAGAATATGAAAATTTTCTATCTTGTGATCTAATATTCA
CAGAATCACAACCTTCCAAAATAATAGGAACTGAAGGAGAATTTTTAGCCTCTAAAAATCTTGATAACAAATCGGGATGCCATGC
AATCATGAACTCTTATATCCATACAAGCAATAATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAAGTAGGATCTTTAACCT
CAAGAGGCGCTGATTCCAATCTTTTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAACTTAACCAGAGAAGAGCATTTAATA
AAAACAAACAAATCATTTAATATATCTATTGATAGCGTTCATGGCATTCATCCAGGATATGCATTTAAACATGATCCAAACTATCA
AGCGACTCTAAGTAAAGGTGTAGTTGTAAAAAATAGCGCAAATTTCAGATACGCAACAACTTCAACAGGATTTGCAAAATTAAAA
AATTTGGCTATTAACAATAATATTGGAAGCGGTAATATTGCTTTTATCCATTTAACCTATGTACCAAGTCCAGCCGGAATTAATGA
GCAAAAATCAAAACCCACTCAACAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTCCCCGATCTGATTATTGCTAGAAGCTCT
CAAGTATTAACAGACCAAATCAGAAAAAAAGTGGCAATGTTTTGCAATGTTGAGAGCACTTCTATTATTGATAATATTGATGTTTC
TACTATTTATGAAATTCCTATATCTTTTTACAAACAAGGCGTACATGAGATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCTAATTTTTTTGTGCCTAAAAAAATTATTAATATTGCCAT
TTGTGGTAAATATGCTGAACTTGATGATTCTTATGCGTCAATTCGAGAGTCTTTGGTTCATGTTGGAGCTAATTTGGACTTACTTAT
TAAAAGTACTCTAATTGATTCTAATGATTTAAATGAGAGTTATTTAAAAGAGTTTGACGGCATTATTGTTCCCGGAGGCTTTGGAG
GCAAAGGATATGAGGGTAAAATTATTGCTATTGCTCGCCAACATTATGATAATTTATCCAGCATATTGTCTTCTTTTGACATTTCAG
TGACTCTTTTGACAGGTAGTTTGAAAAAGAAGGATAAAGAACAAGCATTAGAAAGTATTAAAAGTGGAGTTTCTGGTTTAATAGT
TGGAACACATTCTATTTTTTACGAAAGTACAGAATTTAAAAGATTAGCATATGTTATTATTGACGAGCAGCATAAATTTGGAGTTG
TTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGGGTAGACATACTTTTAATGTCTGCAACACCTATTCCTAGAAGCTTTGCGCT
AACACTTTTTGGTGATCTTGAAGTCTCACTTATTAAAACCTTACCCAAGGGCCGTTTACCTATTACTACTTATTTAGCAAAACATGG
CAATGAAGATAAAGTTTATGATTTTTTAAGAAAAGAGCTCTTAAAGGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTTC
AGAAAAATTCGAATTAAAAGACGTTAATAATATGTATTTAAAATTGAAGGAAGTGTTTAGCGAATATGTTGTAGACATGCTTCAT
TCTAAGTTGCCATCTGATTTAAAAGAAGAAATTATGAAGAATTTTTATTCTAAAAAGGTAGATATTTTGGTGGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTGAAATCGGGCAGAGATTCTTCTGGTAGGATTAGTATTAGAAGAAGAGGT
GGTGGGCATAAGAGAAAATATAGGTTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTGCTTCTATTGAATA
TGATCCTAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAAGGAATTAAGG
TTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGAAATGCTTTACCCCTTGAAAACATTCCTATTGGAAGAACT
ATTCACAATATCGAGCTTAATGTAGGAAAGGGCGGACAGCTGTAAGAAGTGCTGGGGGGTATGCCATGATACTTGCTTCTGATG
GGAATTATGTCACTGTAAAATTGTCATCTGGTGAGATGAGGTTAATTTTCAAAAAATGTATTGCAACAATTGGTGAAATTGGGAAT
GAAGATTATGCTAATGTTTCTATTGGGAAGGCTGGTAAAAGTAGGTGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGACCATCCGCATTCGGGCAGTCTATCTGGGGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGATCAGCACTT
TCAGGCGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAACGAAAAATTAATCTCTACTCTTGTTAATCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAGCACGACGAGCAAACTTTGCGTACTGCAGACTATATTATTGATATGGGTCCTGG
TGCCGGAATTCTTGGAGGAGAAATAGTTGCAAAGGGAGCCTTAATTGATATTTTAAATAGCAAAAACAGTTTAACTGGTCAGTAT
TTGAGCGGCAAGTTTAAAATAGATGTTCCAAGCTCTAGAAGAAAAACAGATAAGGGAGAAATTTTGCTTTTAGGCTCTAATAAAA
ACAATCTTAAAAATATAGATGTAAGTATCCCTTTGGGAGTTTTCACTGTAATAACAGGTGTTTCTGGTAGTGGAAAAAGTACTTTG
CTTAACGAAGTGTTATATCCAGCTCTTGATAGCAGATTAAAGCTTGATGAAAAGTATTGTGATGGGTTTAAAGAC

>BlusitaniaePoHL1
AAAAAAGAACTTCTTATACATGATAGTTTGGTATTTGATTTGATATTAAATATTAAACTTTTAAAATTTAATTTACTTGTCAATAGA
AGTACTATTGGCATATTTGTCTTTATTGGTGCTTCTGGAGTAGGAAAATGCAAATTGACAGATATTTTATCAGAAGAGTTTAAAAT
TCCGAAATTTAGTCTTAATATGAGTGAGTATTGTGATTTTAATTCTCTTGATAGATTAATTGGTCCTATTTTGAATAATGAAGGGCA
TTATGAATCTACTAGATTTTTCAAATTTTTGAACAAATCTTCTAATTCTATTATTTTTCTATCAGATTTTGATAAGTGTAATAAAAG
GGTTTTAGATTTTTTTCTTGAAGGGTTTAAAACAGGCAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCAGAAAGTTTAA
TAGTAATAAGTGTTAATGTTGAGGGCAATGAGCTTAATAGCATCGGCTTTAAAAATAAAATGGCAGAGGAAAATGATTTTAATTT
TATATTAGAGAAGAGATTTCCTAATGAATTTTTAGAGTTAATAGATTATGTGTTTGTATTTgctGGATATGTTGGTGAAGATGTAGA
AAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGTCTAGCTGAGAAGGGAATTATTTATATAGATGAAATAGATA
AAATTGCTAAAAAAAAATGAAAACGTTTCAATAACAAGAGATGTTTCTGGAGAAGGTGTTCAACAGGCTTTGTTAAAGATAATTGA
GGGTACTGTTGCCAATGTTCCCCCAAGAGGTGGGAGAAAACACCCTTATGAGGATACTATTGAAATTAACACTCAAAATATACTC
TTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAATATTGTTAAGAATCGAATAAATAAGAGTTCTATTGGGTTTTCAGCAATTGA
AAAAAAGAGCATAAGAGAAGACACTTCATTAAAATATTTGGAAATGGAAGATTTGATTAAATTTGGTTTAATACCAGAGTTTGTT
GGCAGACTTCCTGTACATTCGTATCTTGAAAAGTTAAATAAAAAAGATTTGATGAAAATATTAGTTGATCCTCAAAATTCTATTGT
TAAGCAGTATTACCATATGTTTAAAATGGATAATGTTGAATTAGTATTTGAAAAAGATGCTTTAGAATCAATTGTAGATGAAGCCC
TTGACAATAAACATATAATTTATTTTGACAATGCAGCAACCTCTCAAAAACCCAAAAAGGTAATTTATTCTAGCATTGAATATTAT
GAAAATTATAACGCAAATGTATACAGAAGCGGTCACAAATTTGCAATTCAATCTAGCATAAAAATAGAAAAAACAAGAGAACTT
GTAAAAAATTTCATTAATGCAGAGTCTACAAAAAATATAATATTTAACTCTGGAACTACAGATGGAATTAATACTATCGCAAACT
CATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACGACTCTTGAACATAATAGTAATTTGCTACCATGGGCA
AATCTTGCAAATTTAGCTAACCTAACAATTAAGTTTGCTAAATTTAATGAAATGGGAATTATTACTCCTGAAGAAATTGAAAAACT
TATTACAGAAAAAACAAAACTCATCAGTATTTCAGGAATAAATAATACCTTAGGAACCATGAATGATTTGGAATCTATTGGAAAA
ATCGCAAAAAAATACAATATAAGTCTTTTTGTAGATGCTGCACAAATGGCGTATAATACTCATGACAATTTAACAGTAATAAACA
GCACTAAAAAAACAATAAAAGATCATATCTTAGAACAACTTGAAATAGAATATGAAAATTTTTTATCTTGTGATCTAATATTCACA

GATTCACAGCCTTCCAAAATAATAGGAACTGAAAAAGAATTTTTAACTTCGAAAAATCTTGATAACAAATCGGGATGCCATGCAA
TCATGAACTCTTATGTTCATACAAGCAATAATAAAAATAAAATAGCTGTATTTTTTGATAACGAAGAAGTAGGATCTTTAACTTCA
AGAGGCGCTGATTCCAATTTTTTATCAGAAGTTTTAGAAAGAATTGATCTTGCTCTTAACTTAACCAAAGAAGAACATTTAATAAA
AATAAACAAATCATTTAATATATCAATCGACAGTGTTCATGGCATTCACCCAGGATATATATCCAAACACGATCCAAACTATCAA
GCAGCTCTAAGCAAAGGCTTAGTTGTAAAAAATAGCGCCAATTTTAAATATGCAACAACTTCAGCAGGATTTGCAAAATTAAAAA
ATTTGGCTATTAAAAATAATATTGGAAGTGGCAATATTGCTTTTATTCATTTAACCTATGTACCAAGCCCGGCTGGAATAAATGAG
CAAAAATCTAAACCCACTCAACAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTTCCCGATTTGATTATTGCTAGAAGTTCTCA
AGTATTGACAGATCAAATCAGAAAAAAAGTGGCAATGTTTTGTAATGTGGAGAGCACTTCTATTATTGATAATGTTGATGTTTCTA
CTATTTATGAAATTCCTATATCTTTTTACAATCAAGGCGTACATAAGATTTTAAGCTCTAAGTTGAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCTAATTTTTCTGTGCCCAAAAAAATTATTAATATTGCTAT
TTGTGGTAAATATGCTGAACTAGATGATTCATATGCGTCAATTAGAGAATCTTTGATCCATGTTGCAGCTCATTTGGATTTACTTAT
TAAAATCACCCTAATTGATTCTAATGATTTAAATGAGGATTGTTTAAAAGAGTTTGATGGCATTATTGTTCCTGGTGGCTTTGGAA
GCAAAGGATATGAGGGTAAAATGATTGCTATTGCTCGTCAACATTATGATAATTTATCTAATATATTGTCGTCTTTTAACATTTCG
GTGACTCTTTTGACTGGTAGTTTAAAAAAGAGGGATAAGGAGCAAGCGTTGGAAAGCATTAGAAGTGGAGTTTCTGGTTTAATAG
TTGGAACACATTCTATGTTTAACGAAAGTACAGAATTTAAAAGATTGGCATATGTTATAATCGACGAGCAGCATAAATTTGGAGT
TATTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGGGTAGATATACTTTTAATGTCTGCAACACCTATTCCTAGAAGTTTTGCG
TTAACACTTTTTGGTGATCTTGAAGTTTCGCTTATTAAAACCTTGCCTAAGGGGCGTTTGCCTATTACTACTTATTTAGCAAAGCAT
GGCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAAGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATC
TTCAGAAAAATTTGAATTAAAAGACGTTAATAATATGTATTTAAAATTGAAGGAAGTTTTTAGCGAATATGTTGTTGACATGCTTC
ATTCTAAGTTGCCATCTGATTTAAAAGAAGAAATCATGAAAAATTTTTATTCTAAAAAGGTGGATATTTTAGTTGCTACTAGTGGT
AATGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCGGGCAGAGATTCTTCTGGTAGGATTAGCGTTAGAAGAAGAG
GTGGTGGGCATAAGAGAAAGTATAGGGTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTGCTTCTATTGAG
TATGATCCTAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGGTATATTATTTCGCCTAAAGGCATTAA
GGTTGGAGATATTTTGGAAAGTGGTCCAAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAA
CTGTTCACAATATTGAGCTTAATGTAGGAAAGGGTGGACAGCTTGTAAGAAGTGCTGGCGGGTATGCTATGATACTTGCTTCTGAT
GGGAATTATGTCACTTTAAAATTGTCTTCTGGCGAGATGAGGTTAATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAAA
TGAAGATTATGCCAATGTTTCTATAGGGAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCTATG
AATCCTGTTGATCATCCGCATTCGGGTAGTCTGTCTGGGGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGCTCAGCACT
TTCAGGTGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAACGAAAAATTAATTTCTACTCTTGTTAATCT
TAAAAATCTTGGCAATACGGTAATTGTTGTTGAACATGATGAGCAAACTTTGCGTACCGCAGACTATATTAATGATGGGCCCTG
GTGCTGGAATTCTTGGAGGAGAAATAATTGCAAAGGGAACCTTAATTGATATTTTAAATAGTGAAAATAGTCTAACTGGTCAATA
TTTGAGCGGTAAGTTTAAAATAGATGTTCCAAGCTCTAGAAGAAAAGCAGATAAGGGAGAAATTTTGCTTTTGGGTTCTAATAAA
AACAATCTTAAAAATATAGACGTAAGGATTCCTTTGGGAGTTTTTACTGTAATAACTGGTGTTTCTGGTAGTGGAAAAAGTACTTT
ACTTAATGAAGTGTTATACCCAGCTCTTGACAGCAGATTAAAGCTTAATGAAAAGTATTGTGATGGTTTTAAAGAT

>BturdiYa501
AAAAAAGAACTTATTATAAATGATAGTTTGGCATCTGATTTAATATTAAATATTAAATTATTAAAATTTAATATACTTGCTAATAG
AAGTACTATTGGTATATTTGCTTTTATTGGTGCTTCTGGAGCCGGAAAATGCAAATTGGTGGATATTTTGTCAGAAGAGCTTAAAA
TTCCAAAATTTACTCTTAATATGAGTGAGTATGGCGATTTTAATTCTCTTGATAGATTAATTGGGCCTGTTTTAAGTAATGAGGGGT
ATTATGAATCTACTAGATTTTTCAAATTTTTGAATAAATTTTCTAATTCTATTGTTTTCCTGTCAGATCTTGATAAATGTAATAAAA
GGGTTTTGGATTTTTTTTTAGAGGGGTTTAAAACAGGTAAACTTTTTGATGGTCTTGGGAGAAAGGCAAGTTTATCAGAAAGTTTA
ATAGTAATAAGTGTCAATGCAGAGAATAATGAGCTTAATAGCATTGGCTTTAAAAATAAAATGGCAGGAGAAAATGATTTTAATC
TTATATTAGAGAAGAGGTTTCCCAATGAATTTTTAGAGTTAATAGATTATGTGTTTTTATTTGCTGGATATGTTGGCGAAGATGTA
GAAAATATTTTACTTAAATTAATACATGCTGCTCATGGAGATGTTAGTTTAGCTGAGAAAGGGATTATTTATATAGATGAAATAGA
TAAAATTGCTAAAAAAAAATGAAAACGTTTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAACAGGCTTTGTTAAAGATAATT
GAGGGGACTATTGCCAATGTTCCTCCAAGAGGTGGAAGAAAACATCCTTATGAGGATACTATTGAAATTAATACCCAAAATATAC
TCTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAATATTGTCAAGAATCGAATAAATAAAAAGTTCTATTGGGTTTTCAGCAATTG
AAAAAAAGAATATAAGAGAGGATACTTCATTAAAGTATTTGGAAATGGAAGACTTAATTAAATTTGGTTTGATACCAGAGTTTGT
TGGTAGACTTCCTGTACATTCGTATCTTGAAAAGTTAAATAAAGAAGATTTGATGAGAATATTAGTTGATCCCCAAAATTCTATTG
TTAAGCAGTATTATCATATGTTTAAAATGGACAATGTGGAGTTAGTATTTGAAAAAGATGCTTTAGAATCAATTGTAGATGAAGCT
TTTGACAATAAGCAAATAATTTATTTTGACAATGCGGCAACCTCTCAAAAGCCTAAAAAAGTAATTTATTCTAGTATTGAATATTA
TGAAAATTGCAACGCAAATGTACACAGAAGTGGTCACAAATTTGCAATTCAATCTAGCATAAAAATAGAAAAAACAAGAGAACT
TGTAAAAAAATTTCATTAATGCAGAATCTACAAAAAATATAATATTTACTTCTGGAACTACAGATGGAATTAATACCATTGCAAGCT
CATTTTTGTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAGCATAATAGTAATTTGCTGCCATGGGCA
AATCTTGCAAATTTAGCTAATTTAACAATTAAATTTGCTAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAACT
TATTACAGAAAAAACAAAACTCATTAGCATTTCAGGAATAAATAATACCTTAGGAACTGTTAATGATTTGGAATCTATTGGGAAA
ATAGCAAAAAAATATAATATAAGTCTTTTTATAGATGCTGCACAAACAGCATATAATACTCATGACAATTTAACAGTAATCAATA
GCACTAAAAAAACAATAAAAGATAATATCTTAGAACAACTTGGAATAGAATATGAAAATTTTCTATCTTGTGATCTAATATTCAC
AGAATCACAACCTTCTAAAATAATAGGAACTGAAGGAGAATTTTTAGCTTCTAAAAATCTTGATAACAAATCAGGATGTCATGCA
ATCATGAACTCTTATATCCATACAAACAATAATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAAATAGGATCTTTAACCTC
AAGAGGCGCTGATTCCAATTTTTTATCAGAAGTTTTAGAAAGAATTGATCTTGCTCTTAACTTAACCAGAGAAGAGCATTTAATAA
AAATAAACAAATCATTTAATATATCGATCGACAGCGTTCACGGTATTCATCCAGGATATGCATTCAAACATGATCCAAACTATCA
AGCAACTTTAAGTAAAGGCGTAGTTGTAAAAAATAGCGCCAATTTCAGATACGCAACAACTTCAAAAGGATTTGCAAAATTAAAA
AATTTGGCTATTAAAAGTAATATTGGAAGTTGCAATATTTCTTTTATTCATTTAACTTATGTGCCAAGTCCGGTCGGAATTAATGAG
CAAAAATCTAAACCTACTCAACAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTTCCGGATCTGATTATTGCTAGAAGTTCTCA
AGTATTAACAGATCAAATCAGAAAAAAAGTAGCAATGTTTTGTAATGTTGAGAGCGCTTGTATTATTGATAATGTTGATGTTTCTA
GTATTTATGAAATTCCTATATCTTTTTACAAACAAGGTGTACATGAGATTTTAAGTTCTAAGTTGAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCTAATTTTTTTGTGCCTAAAAAAATTATTAATATTGCTAT
TTGTGGTAAATATGCTGAACTTGATGATTCTTATGCGTCAATTAGAGAGTCTTTGGTCCATGTTGGAGCTAATTTAGATTTACTTAT
TAAAAGTGCCCTAATTGATTCCAATGATTTAAATGAAGCTTGTTTAAAGCAGTTTGATGGCATTATTGTTCCTGGCGGTTTTGGAG
GGAAAGGATATGAAGGTAAAATTATTGCTATTGCTCGTCAACATTATGATAATTTATCCAACATATTGTCGTCTTTTAACGTTTCA
GTGACTCTTTTAACCGGTAGTTTAAAAAAAAGAGATAAGGAACAAGTGCTGGAAAATATTAAAAGTGGAACTTCTGGTTTAATAG
TTGGAACTCATTCTATTTTTTACGAAAGTACAGAATTTAAAAAATTAGCATATGTTATCATTGACGAGCAGCATAAATTTGGAGTT
GCTCAAAGAGAAGAACTTAAAAATAAAGGAGAAGGGGTAGATATGCTTCTAATGTCTGCAACACCTATTCCTAGAAGCTTTGCGT
TAACACTTTTTGGCGATCTTGAAGTTTCGTTTATTAAAACCTTGCCTAAGGGGCGTTTGCCTATTACTACCTATTTAGCAAAACATG
GTAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTTTCAAAGGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTT
CGGAAAAATTTGAATTAAAAGATGTTAATAATATGTATTTAAAATTGAAGGAAGTATTTAGCGAATATGTTGTTGAGATGCTTCAT
TCTAAGTTACCATCTGATTTAAAAGAAGAAATTATGAAAAATTTTTATTCTAAAAAAGTAGATATTTTGGTGGCTACTAGCGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAGTTTAAATCAGGTAGAGATTCTTCTGGTAGGATTAGTATTAGAAGAAGAGGT
GGTGGGCATAAGAGAAAGTATAGGGTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCCGCTCGAGTTGCTTCTATTGAAT
ATGATCCTAATAGAAGTGCCAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAAGGCATTAAG
GTTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAAC
TGTTCATAATATTGAGCTTAATGTGGGAAAAGGTGGGCAGCTTGTAAGAAGTGCTGGTGGATATGCCATGATACTTGCTTCTGATG
GGAACTATGTCACTGTAAAATTGTCATCTGGCGAGATGAGGTTAATTTTCAAAAAATGTATTGCGACAATTGGTGAGATTGGGAA
TGAAGATTATGCCAATATTTCTATAGGGAAAGCTGGTAAAAGTAGGTGGCTTGGGAGAAGGCCAAAGGTTAGGGGTGTTGCCATG
AATCCCGTTGACCATCCGCATTCAGGCAGTCTATCTGGAGGTGAAGCTCAGCGCATTAGGCTTGCTACTCAAATAGGGTCAGCACT

TTCGGGTGTTATTTATGTTCTTGATGAACCAAGTATCGGGCTTCATCAAAAAGATAATGAAAAATTAATCTCTACTCTTGTTAATCT
TAAAAATCTTGGCAATACAGTAATTGTTGTTGAGCATGACGAGCAAACTTTGCGTACCGCAGACTATATTATTGATATGGGCCCCG
GTGCTGGAATTCTTGGGGGGAAAATAGTTGCAAAGGGAGCTTTAATTGATATTTTAAATAGTAAAAATAGTTTAACCGGGCAATA
TTTAAGCGGTAAGTTTAAAATAGATGTTCCAGACTCTAGAAGAAAGACAAATAAGGGAGAAATTTTGCTTTTAGGAGCTAATAAA
AACAATCTTAAAAATATAGATGTAAGCATCCCTTTAGGAGTTTTTACTGTAATAACAGGTGTTTCTGGTAGCGGAAAAAGTACTTT
GCTTAACGAAGTACTTTTTCTAGCTCTTGATAGCAGATTAAAGCTTGATGAAAAGTATTGTGATGGTTTTAAAGAT

>BspeilmaniiA14S
AAAAAAGAACTTATTATAGATGATAGTTTAGTATTTGATTTAATATTAAATTTAAAATTATTAAAATTTAATTTACTTACCAATAG
AAGCACTATTGGTATATTTGCCTTTATTGGTGCTTCTGGAGTGGGAAAATGCAAATTGACTGATATCTTATCAGAAGAGCTTAAAA
TTCCGAAATTTAATATTAATATGAGTGAGTATAGTGATTTTAATTCTTTTGATAGATTAATTGGTCCTGTTTTAAGTAATGATGGGT
ATTATGAATCTACTAGATTTTTCAAATTTTTGAATAAATCTTCTAATTCTATTATTTTCCTGTCAGATTTTGATAAATGTAATAAAA
GGGTTTTAGAATTTTTTTTAGAAGGGTTTAAAACTGGCAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGCTTATCGGAAAGTTTA
ATAGTAATAAGCGTCACTGTTGAGAGTAATGAGCTTAATAGTATCGGCTTTAAAAATAAATCGACAGGTGGAAATGATTTTAATC
TTATATTAGAGAAGAGATTTTCTAATGAATTTTTAGAGTTAATAGATTGTATGTTTGTATTTGCTGGATATGTTGGTGAGGATGTAG
AAAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGTCTAGCTGAGAAAGGGATCATTTATATAGATGAAATAGAT
AAAATTGCTAAAAAAAAATGAAAACGTTTCAATAACAAGAGATGTATCTGGAGAAGGCGTTCAGCAAGCTTTGTTAAAGATAATTG
AGGGTACTGTTGCTAATGTTCCTCCAAGAGGGGGTAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATACT
CTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAAATATTGTTAAGAATCGAATAAATAAGAGTTCTATTGGATTTTCAGCAATTG
AAAAAAAGAATATAAGAGAGGACACTTCATTAAAGTATTTAGAAATGGAAGATTTGATTAAATTCGGTTTGATACCAGAGTTTGT
TGGCAGACTTCCTGTACATTCGTATCTTGAAAAATTAAGCAAAGAAGATTTGATGAGAATATTAGTTGATCCTCAAAATTCTATTG
TTAAGCAGTATTATTATATGTTTAAAATGGATAATGTTGAATTGGTATTTGAAAAAGATGCATTAGAATCAATAGTAGATGAAGCT
TTTAATAATAAGCCTATAATTTATTTTGACAACGCAGCAACCTCTCAAAAACCCCAAAAAGTAATTTACTCCAGCATTGAATATTA
TGAAAATTACAACGCAAATGTACATAGAAGCGGTCACAAATTTGCAATCCAATCTAGCATAAAAATAGAAAAAACAAGAGAACT
TGTAAAAAACTTCATTAATGCAGAATCTGCAAAAAATATAATATTTACTTCTGGAACTACAGATGGAATTAATACCATTGCAAACT
CATTTTTTTGCTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAATAGCAATTTGCTGCCATGGGCA
AATCTTGCAAAATTAGTTAATCTAACAATTAAATTCGCTAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAAC
TTATTACAGAAAAAACAAAGCTCATCAGTATTTCAGGAACAAATAATACCTTGGGAACCGTCAATGATTTGGAATCTATTGGAAA
AATCGCAAAAAAATACAATATAAGTCTTTTTGTAGATGCTGCCCAAATGGCATATAATACTCACGACAACTTAACAGTAATCAAC
AGTACTAAAAAAACAATAAAAGATAATATCTTAGAACAACTTGGAATAAAATATGAAAATTTCTATCCTGTGATCTAATATTTA
CAGAATCACAACCTTCTAAAATCATAGGAACTGAAGGAGAATTTTTAACTTCTAAAAATCTCGATAACAAATCGGGATGCCACGC
AATCATGAATTCTTATGTTCATACAAATAATAATAAAAATAAAATAGCTGTATTTTTTGATAACGAAGAAGTAGGATCTTTAACCT
CAAGAGGTGCTGACTCTAATTTTTTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAACTTAACCAGAGAAGAGCATTTAATA
AAAACAAACAAATCATTTAATATATCGATTGACAGCGTTCATGGCATTCATCCAGGATATGCATTTAAACATGACCCAAACTATC
AAGCAACTCTAAGCAAAGGCGTAGTTGTCAAAAATAGCGCCAATTTTAGATATGCAACAACTTCAAAGGGATTTGCAAGATTAAA
AAATTTAGCTAATGAAAATAATATTGGGAGTGGTAATATTGCTTTTATTCATTTAACTTATGTGCCAAGTCCAGCAGGAATTAATG
AGCAAAAATCCAAACCTACTCAACAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTCCCCGATTTAATTATTGCAAGAAGCTC
TCAAGTATTGACAGATCAAATCAAAAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGTACTTCTATTATTGATAATGTTGATGTTT
CTACTATTTATGAAATTCCCATATCTTTTTACAAACAAGGTGTGCATGAGATTTTAAGTTCTAAGCTAAATATTAGGGTT---
GATCCAAAAATAGAAGAGCTTTCAAGGCTTGTAGGAGTTATAAAATCTAATTTTTTTATGCCTAAAAAAATTATTAATATTGCTAT
TTGCGGTAAATATGCTGAACTTGATGATTCTTATGCATCGATTAGAGAGTCTTTGGTTCATGTTGCAGCTAATTTGGACTTACTTAT
TAAAAGCACCATAATTGATTCTAATGATTTAAATGAGAGTCGTTTGAAAGAGTTTGACGGCATTATTGTTCCTGGTGGTTTTGGAG
GTAAAGGATATGAAGGCAAAATTACTGCTATTGCTCGTCAACATTATGATAATTTATCCAATATATTGTCCCCTTTAAACATTTCA
GTGACTCTTTTGACCGGTAGTTTAAAAAAGAGGGATAAAGAACAAGCGTTGGAAAGTATTAGAAGTGGAACTTCTGGTTTGATAG
TTGGAACACATGCTATTTTTTACGAAAGCACAGAATTTAAGAGATTAGCATATGTTATTATTGATGAGCAACATAAATTTGGAGTT
TTTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGAGTAGACATGCTTTTAATGTCTGCAACTCCTATTCCTAGAAGTTTTGCTTT
AACACTTTTTGGTGATCTTGAAGTTTCTTTTATTAAAACCTTGCCTAAGGGGCGTTTGCCTATTACTACTTATTTAGCAAAGCATGG
CAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAGGGTCATCAGGTATATTTTGTTTATCCATTAATTTCATCTTC
AGAAAAATTTGAATTAAAAGACGTTAATAATATGTGTTTAAAATTAAAAGAAGTGTTTAGCGAATATGTTGTCGATATGCTTCATT
CTAAGTTGCCGGCTGATTTAAAAGAAGAAATTATGAAGAATTTTTATTCTAAAAAAGTAGATATTTTGGTGGCTACTAGTGGTAAT
GATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAGTCGGGCAGAGATTCTTCTGGTAGGATTAGTGTTAGAAGAAGAGGTG
GTGGGCATAAGAGAAGGTATAGGGTAATAGATTTTAATCGAAGAGATAAATTTAGCATTCCCGCTCGAGTTGCTTCTATTGAATA
TGATCCTAATAGAAGCGCTAATATGCTTGCTTGTTTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAAGGCATTAAGG
TTGGTGATGTTTTGGAAAGTGGTCCGAATTCACCAATTAAAATTGGTAATGCTTTGCCTCTTGAAAATATTCCTATTGGAAGAACT
GTTCATAATATAGAGCTTAATGTGGGAAAGGGTGGACAGCTTGTAAGAAGTGCTGGGGGATATGCTATGATACTTGCTTCTGACG
GGAATTATGTCACTGTAAAATTATCGTCTGGTGAAGTGAGGTTGATTTTCAAAAAATGCATTGCAACAATTGGTGAAATTGGGAA
TGAAGATTATGTCAATGTTTCTATAGGAAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCCATG
AATCCTGTTGACCATCCGCATTCTGGCAGTCTGTCTGGGGGTGAGGCCCAGCGTATTAGGCTTGCTACTCAAATAGGGTCGGCGCT
TTCGGGTGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAATGAAAAATTAATTTCTACTCTTGTTAATCT
TAAAAATATTGGCAATACAGTAATTGTTGTTGAACATGATGAACAAACTTTGCGTACTGCAGACTATATTATTGATATGGGGCCTG
GTGCTGGAATTCTTGGAGGGGAAATAGTTGCAAAGGGAACTTTAATAGATATTTTAAATAGTAAAAACAGTTTAACCGGTCAATA
TTTGAGTGGCAAGTTTAAAATAGATGTTCCAAGTTCTAGAAGAAAGACAGATAAGGGAGAGATTTTACTATTAGGTTCTAGTAAA
AACAATCTTAAAAATATAGACGTAAGTATTCCTTTGGGAGTTTTTACTGTAATAACAGGTGTTTCTGGTAGTGGAAAAAGTACTCT
ACTTAACGAGGTGTTATATCCAGCGCTTAACAGTAGATTAAAGTCTAATGGAAAGTATTGTGATGGTTTTAAAGAT

>Genomospecies2CA2
AAAAAAGAACTTATTATACATGATAGCTTGATATTTGATTTGATATTAAATATCAAATTATTAAAATTCAATTTACTTGCCAATAG
AAGTACTATTGGCATATTTGCCTTTATTGGTGCTTCTGGATCAGGAAAATGCAAATTGACGGATATTTTATCAGAAGAGTTTAAAA
TTCCGAAATTTACTCTTAACATGGGTGAGTATAGTGATTTTAATTCTCTTGATAGATTGATTGGGCCTGTTTTAAGTAATGAAGGAT
ATTATGAATCTACTAGATTTTTTAAATTTTTAAACAAATCTTCTAATTCTATTATTTTTTTATCAGATTTTGATAAATGTAATAAAAA
GGTTTTAGATTTTTTTTTGGAGGGGTTTAAAACAGGTAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCGGAAAGTTTAA
TAGTAATAAGTATCAATGCTGAAAGCAAAGAGCTTAATAGCATTGGTTTTAAGAATAAAATGGCAGGGGGGGGT---
TTTAACCTTATATTAGAGAAGAGATTTCCTAATGAGTTTTTAGAGTTAATAGATCATGTGTTTGTATTTGCAGGATATGTTGGTGAA
GATGTGGAAAATATTTTACTTAAATTGATACATGCTGCTCATGGGGATGTTAGCTTGGCTGAGAAAGGGATTATTTATATAGACGA
GATAGATAAAATTGCTAAAAAAAATGAAAACGTATCAATAACAAGAGATGTTTCTGGGGAAGGGGTTCAACAGGCTTTGTTAAA
GATAATTGAAGGTACTGTTGCCAATGTTCCTCCAAGAGGCGGTAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAA
AATATACTCTTTATATGTGGTGGTGCTTTTGTTGGTCTTGAAAATATTGTTAAGAATCGGATAAATAAAAGTTCTATTGGGTTTTCA
GCAATTGAAAAAAAGAATATAAGAGAAGATACTTCATTAAAGTATTTAGAAATGGAAGATTTGATTAAATTTGGCTTAATACCAG
AGTTTGTTGGCAGACTTCCTGTGCATTCGTATCTTGAAAAGTTAAATAAAGAAGATTTGTTGAAAATATTAGTTGATCCTCAAAAT
TCTATTGTTAAGCAGTATTATCATATGTTTAAAATGGACAATGTTGAATTGGTATTTGAAAAAGATGCTTTAGAATCAATTGTAGA
TGAAGCCTTTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACCTCTCAAAAGCCCAAAAACGTAATTTATTCTAACGTTA
AATATTATGAAAATTATAACGCAAATGTACACAGAAGCGGTCATAAATTTGCAATTCAATCGAGCATAAAAATAGAAAAAACAA
GAGAACTTGTAAAAAATTTCATCAATGCAGAATCTGCAAAAAATATAATATTCACCTCTGGAACTACTGATGGAATTAATACTATT
GCAAGTTCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAATAGTAATTTGCTGCC
TTGGGTAAATCTTGCAAATTTAGCTAATTTAACAATTAAATTAGCTAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTG
AAAAACTTATTACAGAAAAAACAAAGCTCATCAGTATTTCAGGAACAAATAATACCTTAGGAACCATTAATGATTTGGAATCTAT

TGGAAAAATCGCAAAAAAATACAATATACCTCTTTTTGTAGATGCTGCACAAATGGCATACAATACTCATGACAATTTAACAGTA
ATCAGCAGCACTAAAAAAACAATAAAAGATAATGTCTTAGAACAACTTGGAATAGAGTGTGAAAATTTTTTATCTTGTGATTTAA
TATTTACAGAATCACAACCTTCCAAAATAATAGGAACTGAAGGAGAATTTTTGGCTTCTAAAAATCTTGATAACAAATCGGGATG
CCATGCAATCATGAATTCTTATGTTCATACAAGCAATGATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAAGTAGGATCTT
TAACCTCAAGAGGTGCTGATTCAAACTTTCTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACCAGAGAAGAGCAT
TTAATAAAAACAAACAAATCATTTAATATCTCGATTGACAGCGTTCACGGCATTCATCCGGGTTATACATCTAAACATGATCCAAA
CTATCAATCAACTCTGGGTAAAGGCGTAGTTGTAAAAAATAGCGCCAATTTCAGATATGCAACAACTTCAACAGGATTTGCAAAA
TTAAAAAATTTGGCTATTAAAAATAATATTGGAAGTGGTAATATTGCCTTTATTCATTTAACCTATGTACCAAGCCCAGCTGGAAT
TAATGAGCAAAAGTCTAAACCTACTCAACAGAGTGTTAAAACTTTAAATAAAGCGGGTATTTTTCCTGATTTAATTATTGCTAGAA
GTTCTCGGGTATTGACAGATCAAATCAGAAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGCACTTCTATTATTGACAATGTTGAT
GTTTCTACTATTTATGAGATTCCTATATCTTTTTACAAACAAGGCGTACATGATATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAGGAGCTTTCAAAGCTTGTAGGAATTATAAAATCTAATTTTTTTGCGCCTAAAAAAATTATTAATATTGCTAT
TTGTGGTAAATATGCTGAGCTTGATGATTCTTATGCGTCAATTAGAGAGTCTTTGGTTCATGTTGCAGCTCATTTGGACTTGCTTAT
TAAAAGTACTTTAATTGATTCTAATGATTTAAATGAGGTCTGTTTAAAAGATTTTGATGGCATTATTGTTCCTGGCGGCTTTGGGGG
AAAAGGATATGAAGGCAAAATTATGGCTATTGCTCGTCAACATTATGATAATTTATCTAACATATTGTCGTCTTTTAACATTTCGG
TGACTCTTTTGACTGGTAGTTTAAAAAAGAAGGATAAGGAGCAAGCATTAGAAAGTATTAAAAATGGAAATTCTGGTTTAATAGT
TGGAACACATGCTATTTTTTGCGAAAGTACAGAATTTAAAAGATTAGCATATGTTATCATTGACGAGCAGCATAAATTTGGAGTTG
TTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGGGTAGACATGCTTTTAATGTCTGCAACGCCTATTCCCAGAAGTTTTGCGTT
AACACTTTTTGGTGATCTTGAAATTTCGTTTATTAAAACCTTACCTAAGGGACGTTTGCCTATTACTACTTATTTAGCAAAGCATGG
CAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAAGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTTC
AGAAAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTAAAAGAAGTGTTTAGCGAGTATGTTGTGGATATGCTTCATT
CCAAGTTGCCATCCGATTTAAAAGAAGAAATTATGAAAAATTTTTATTCGAAAAAAGTAGATATTTTGGTGGCTACTAGTGGCAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCGGGCAGAGATTCTTCTGGTAGGATTAGTATTAGAAGAAGAGGT
GGTGGGCATAAGAGAAAGTATAGGTTGATTGATTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTGCTTCTATTGAATA
TGATCCTAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAAGGCATTAAGG
TTGGAGATGTTTTGGAAAGTGGTTCGAATGTACCAATTAAAATTGGAAATGCCTTACCTCTTGAAAATATTCCTATTGGAAGAACT
GTTCACAATATTGAGCTTAGTGTGGGAAAGGGCGGACAACTTATAAGAAGTGCTGGTGGATATGCTATTATACTTGCTTCTGACG
GGAATTATGTCACTGTAAAATTATCATCTGGTGAGATGAGATTGATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAAAT
GAAGATTATGCCAATATTTCTATAGGGAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGGCCCAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGACCATCCACATTCAGGCAGCCTATCTGGTGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGATCAGCACTT
TCGGGTGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAACGAAAAATTAATTTCTACTCTTGTTAATCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAGCACGATGAGCAAACTTTGCGTACTGCAGACTATATTATTGATATGGGTCCTGG
TGCTGGCATTCTTGGGGGGGAAATAGTTGCAAAGGGATCCTTAATTGATATTTTAAATAGCAAAAATAGCTTAACTGGTCAATATC
TTAGCGGCAAGTTTAAAATAGAAGTTCCAAGCTCTAGAAGAAAGTCGGATAAGGGAGAAATTTTGCTTTTGGGTTCTAATAAAAA
CAATCTTAAAAATATAGACGTAAGTATCCCTTTGGGAGTTTTTACCGTAATAACAGGTGTTTCTGGTAGTGGGAAAAGTACTTTGC
TTAACGAAGTGTTATATCCAGCTCTTGATAGTAGATTGAAGCTTAATGAAAAGTATTGTGATGGTTTTAGAGAT

>Bamericana
AAAAAAGAACTTATTATACAAGATAACTTGGTATTTGATTTGATATTAAATATTAAATTATTAAAATTCAATTTACTTGTCAACAG
AAGTACTATTGGCATATTTGCCTTTGTTGGTGCTTCTGGATCAGGAAAATGCAAATTGACGGATATTTTATCACAAGAGTTTAAAA
TTCCGAAATTTAGTCTTAACATGGGTGGGTATAGTGATTTTAATTCTCTTGATAGATTGATTGGGCCTGTTTTAAGTAATGAAGGGT
ATTATGAATCTACTGGATTTTTTAAATTTTTGAACAAATCTTCCAATTCTATTATTTTTTTATCAGATTTTGATAAATGTAATAAAA
GGGTTTTGGATTTTTTTTTAGAGGGGTTTAAAACAGGTAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCAGAAAGTTTA
ATAGTAGTAAGTATTAATGCTGAGGTCAAGGAGCTTAATAGTATTGGCTTTAAAAATAAAATGGCGGGGGTAAATGATTTTAACT
TTATATTAGAGAAGAGATTGCCTAATGAATTTTTAGAGTTAATAGATCATGTGTTTGTATTTGCAGGATATGTTGGTGAAGATGTA
GAAAACATTTTACTTAAATTGATACATGCTGCTCATGGGGATGTTAGCTTAGCTGAGAAAGGGATTATTTATATAGACGAGATAG
ATAAAATTGCCAAAAAAAACGAAAATGTCTCAATAACAAGAGATGTTTCTGGAGAAGGAGTTCAACAGGCTTTATTAAAGATAAT
TGAAGGCACTATTGCTAATGTTCCCCCAAGAGGCGGTAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACACAAAATATA
CTTTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAATATTATTAAGAATCGAATAAATAGAAGTTCTATTGGGTTTTCAGCAATT
GAAAAAAAGAATATAAGAGAAGATACTTCATTAAAGTATTTAGAAATGGAAGATTTGATTAAATTTGGTTTAATACCAGAGTTTG
TTGGCAGACTTCCTGTGCATTCGTATCTTGAAAAGTTGAATAAAGAAGATTTGTTGAGAATATTGGTTGATCCTCAAAATTCTATT
GTTAAGCAGTACTATCATATGTTTAAAATGGACAATGTTGAATTGGTATTTGAAAAAGATGCTTTAGAATCAATAGTAGATGAGG
CCTTTGAAAATAAACATATAATTTATTTTGACAATGCAGCAACCTCTCAAAAGCCCAAAAACGTAATTTATTCTAACGTTGAATAT
TATGAAAATTACAACGCAAATGTACACAGAAGCGGTCATAAATTTGCAATCCAATCGAGCATAAAAATAGAAAAAACAAGAGAA
CTTGTAAAAAATTTCATTAATGCAGAATCCGCAAAAAATATAATATTCACCTCTGGAACTACTGATGGAATTAATACTATTGCAAG
TTCATTTTTTTACTCAAAATATTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAGCATAACAGCAATTTACTGCCTTGGGT
AAATCTTGCAAATTTAGCTAATTTAAAAATTAAATTAGCTAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAA
CTTATTACAGAAAAAACAAAACTTATCAGTATTTCAGGAATAAATAATACCTTAGGAACCGTTAATGATTTGGAATCTATTGGAA
AAATCGCAAAAAAATACAATATATGTTTTCTTGTAGATGCTGCACAAATGGCATACAATACTCATGACAATTTAACAGTAATAAG
CAGCACTAAAAAAACAATAAAAGATACTGTCTTAGAACAACTTGGAATAGAATGTAAAAATTTTTTATCTTGTGATTTAATATTTA
CAGAATCACAACCCTCCAAAATAATAGGAACAGAAGGAGAATTTATGGCTTCTAAAAATCTTGATAACAAATCGGGATGCCACGC
AATCATGAATGCTTATGTTCATACAAACAATGATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAAGTAGGATCTTTAACCT
CAAGAGGCGCTGATTCAAATTTTCTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACTAGAGAAGAGCATTTAATA
AAAACAAACAAATCATTTAATATCTCGATTGACAGCGTTCACGGCATTCATCCAGGCTATACATCTAAACATGATCCAAACTATCA
AGCAACTTTAGGTAAAGGCCCAGTTGTAAAAACTAGCGCTAATTGCAGATATGCAACAACTTCAACAGGATTTGCAAAATTAAAA
AATTTGGCTATTAAAAATAATATTGGAAGTGGCAATATTTCTTTTATTCATTTAACCTATGTGCCAAGTCCAGCTGGAATTAATGA
GCAAAAGTCTAAACCTACCCAACAGAGTGTTAAAACCTTAAATAAAGCGGGCATTTTCCCCGATTTAATTATTGCTAGAAGTTCGC
AAGTATTAACAGATCAAATCAGAAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGCACCTCTATTATTGACAATGTTGATGTTTCT
ACTATTTATGAAATTCCCATATCTTTTTACAAGCAAGGTGTACATGAAATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCCAATTTTTTTGCGCCTAAAAAAATTATTAATATTGCTGT
TTGTGGTAAATATGCTGAACTTGATGATTCTTATGCATCAATTAGAGAGTCCTTGGTGCATGTTGCAGCCCATTTGGACTTGCTTAT
TAAAAGCACTTTGATTGATTCTAATGATTTAAATGAGAGCTGTTTAAAAGAGTTTGATGGCATTATTGTTCCTGGCGGCTTTGGAG
GTAAGGGGTATGAAGGTAAAATTATGGCTATTGCTCGTCAGCATTATGATAATTTATCTAACATACTGTCGTCTTTTAACATTTCA
ATAACTCTTTTGACTGGCAGTTTAAAAAAGAAGGATAAGGAGCAAGCATTAGAAAGCATTAAAAGTGGCGCTTCTGGCTTAATAG
TTGGAACACATGCTATTTTTTATGAAAGTACAGAATTTAAAAGATTAGCATATGTTATTATTGATGAGCAGCATAAATTTGGAGTT
GTTCAAAGAGAAGAGCTTAAGAACAAAGGAGAGGGAGTAGACATGCTTTTAATGTCTGCAACGCCTATTCCTAGAAGTTTTGCGT
TAACACTTTTTGGTGATCTTGAAGTTTCTTTTATTAAAACCTTGCCTAAGGGTCGATTGCCTATTACTACTTATTTAGCAAGGCATG
GCAATGAAGATAAAGTTTATGAATTTTTAAGAAAAGAGCTCTTAAAAGGGCATCAGGTTTATTTTGTTTATCCATTAATTTCATCT
TCAGAAAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTAAAAGAAGTATTTAGCGAATATGTTGTAGACATGCTTCA
TTCCAAGTTGTCATCCGATTTAAAAGAAGAAATTATGAAAAATTTTTATTCAAAAAAAGTAGATATTTTGGTAACTACTAGTGGTA
ATGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTGAAATCAGGCAGAGATTCTTCTGGTAGGATTAGTATTAGAAGAAGAGG
TGGTGGGCATAAGAGAAAGTATAGGTTGATTGATTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTGCTTCTATTGAAT
ATGATCCCAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAAGGCATTAAG
GTTGGAGATGTTTTGGAAAGTGGTCCTAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAAC
TGTTCACAATATTGAGCTTAATGTTGGAAAAGGTGGACAACTTATAAGAAGTGCTGGTGGATATGCTATGATACTTGCTTCTGACG
GGAATTATGTCACTGTAAAATTATCATCTGGTGAGATGAGGTTGATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAAAT

GAAGATTATGCCAATATTTCTATAGGGAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGACCCAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGACCATCCGCATTCAGGCAGCTTATCTGGTGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGATCAGCACTT
TCGGGTGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAATGAAAAATTAATCTCTACTCTTGTTAATCTT
AAAAATCTTGGTAATACTGTAATTGTCGTTGAGCATGATGAGCAAACCTTGCGTACGGCAGACTATATTATCGATATGGGGCCCG
GTGCTGGCATTCTTGGGGGGGAAATAGTTGCAAAGGGAGCTTTAATTGATATTTTAAATAGCGAAAATAGCTTAACCGGTCAATA
TCTTACCGGCAAGTTTAAAATAGATGTTCCAAGCTCTAGAAGAAGGGCAGATAAGGGAGAAATTTTACTTTTAGGTTCTAACAAA
AATAATCTTAAAAATATAGATGTAAGTATTCCTTTGGGAGTTTTTACCGTAATAACAGGTGTTTCTGGTAGTGGGAAAAGTACTTT
ACTCAACGAAGTGTTATATCCAGCTCTTGATAGTAGATTAAAGCTTAATGAAAAGTATTGTGATGGTTTTAAAGAT

>Bandersonii
AAAAAAGAACTTATTATGCATGATAACTTGGTATTTGACTTGATATTAAATATTAAATTATTAAAATGCAATTTGCTTGCCAACAG
AAGTACTATTGGCATATTTGTTTTTATTGGTGCTTCTGGATCAGGAAAATGTAAGTTGACGGATATTTTATCACAAGAGTTTAAAA
TTCCGAAATTTAGTCTTAATATGGGTGAATATAGTGATTTTAATTCTCTGGATAGATTGATTGGCCCTGTTTTAAGTAATGAAGGGT
ATTATGAATCTACTAGATTTTTTAAATTTTTGAACAAGTCTTCTAATTCTATTATTTTTTTATCAGATTTTGATAAATGTAATAAAAA
GGTTTTAGATTTTTTTTTAGAGGGGTTTAGAACTGGTAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCAGAAAGTTTAA
TAATAATAAGTGTTAATGCGGAGAGAAAGGAGCTTAATAGTATTGGCTTTAAAAATAAAATGGCTGGGGGGAATGATTTTAGTTA
TATTTTAGAAAAAAGATTGCCCACTGAGTTTTTAGAGTTAATAGATCATGTTTTTGTATTTGCAGGATATGTTGGCGAAGATGTAG
AAAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGTTTAGCTGAGAAAGGGATTATTTATATAGACGAGATAGAT
AAAATTGCTAAAAAAAATGAAAATGTTTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAACAGGCTTTGTTAAAGATAATTG
AAGGCACTGTTGCCAATGTTCCTCCAAGAGGTGGTAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATACTT
TTTATATGTGGTGGCGCTTTTGTTGGGCTTGAAAATATTGTTAAGAATCGAATAAATAAAAGTTCTATTGGCTTTTCAGTAATTGA
AAAAAAGAATATAAGAGAAGATACTTCATTAAAGTATTTAGAAATGGAAGATTTGATTAAATTTGGCTTAATACCAGAGTTTGTT
GGCAGGCTTCCTGTGCATTCGTATCTTGAAAAGTTGAATAAAGAAGATTTGTTGAAAATATTAGTTGATCCCCAAAATTCTATTGT
TAAGCAGTATTATCATATGTTTAAAATGGACAATGTTGAATTAGTGTTTGAAAAAGATGCTTTAGAATCAATTGTAGATGAAGCTT
TTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACCTCTCAAAAGCCCAAAAACGTAATTTATTCCAACGCTGAATATTAT
GAAAATTACAACGCAAATGTACATAGAAGCGGTCACAAATTTGCAATTCAATCAAGCATAAAAATAGAAAAAACAAGAGAACTT
GTAAAAAATTTCATTAATGCAGAATCTGCAAAAAATATAATATTTACCTCTGGAACTACTGATGGAATCAATACTATTGCAAGTTC
ATTTTTTTGCTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAACAGTAATTTACTACCTTGGGTAA
ATCTCGCAAATTTAGCTGATTTAAAAATTAAACTAGCCAAATTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAACT
TATTACAGAAAAAACAAAGCTCATCAGTATTTCAGGAATAAATAATACCCTAGGAACCATTAATGATTTAGAATCTATTGGAAAA
ATTGCAAAAAAATACAACATATGTCTTTTTGTAGATGCTGCACAAATGGCATACAATACTCATGACAATTTAACAGTAATCAGTA
GCACTAAAAAAACTATAAAAGATAATATCTTAGAACAACTTGGAATAGAACATAAAAATTTTTTATCTTGTGATTTAATATTTACA
GAATCACAACCTTCTAAAATAATAGGAACAGAAGGAGAATTTTTGGCTTCTAAAAATCTTGATAACAAATCGGGATGCCATGCAA
TCATGAATTCTTATGTTCATACCAGCAATGACAAAAATAAAATAGCTGTATTTTTTGATAACGAAGAAGTGGGATCTTTAACCTCA
AGAGGAGCTGATTCAAATTTTCTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACCAAAGAAGAGCATTTAATAAA
AATAAACAAATCATTTAATATTTCGATTGACAGCGTTCACGGCATTCATCCGGGATACACATCTAAACATGATCCAAACTATCAAG
CTACTCTGGGCAAAGGTGTAGTTGTAAAAACTAGCGCTAATTTCAGATATGCAACAACTTCAACAGGATTCGCAAAATTAAAAAA
TTTGGCTATTAAAAATAATATTGGAAGTGGCAATATTTCTTTTATTCATTTAACCTATGTACCAAGTCCAGCTGGAATTAATGAAC
AAAAATCTAAACCTACTCAACAGAGTGTTAAAACCTTAAACAAAGCAGGTATTTTCCCTGATTTAATTATTGCCAGAAGTTCTCAA
GTATTGACAGATCAAATTAGAAAAAAAGTGGCAATGTTTTGTAATGTTGATAGCGCTTCTATTATTGACAATGTTGATGTTTCTAC
TATTTATGAAATTCCTATATCTTTTTACGAGCAAGGTGTACATGAGATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAAGAACTTTCAAAGCTTGTGGGAGTTATAAAATCCAATTTTTTTGCGCCTAAAAAAATTATTAATATTGCTGT
TTGTGGCAAATATGCTGAACTTGATGATTCTTATGCATCAATTAGAGAGTCCTTGGTTCATGTTGCAGCTCATTTGGACTTACTTAT
TAAAAGTACTTTAATTGATTCTAATGATTTAAATGAGGCCTGTTTAAAAGAGTTTGACGGTATTATTGTTCCTGGTGGCTTTGGAG
GTAAAGGATATGAGGGTAAAATTATGGCTATTGCTCGTCAACATTATGATAATTTATCTAAGATGTTGTCGTCTTTTAACATTTCG
GTTACTCTTTTGACTGGCAGTTTGAAAAAGAGGGATAAGGAGCAAGCATTAGAAAGTATTAAAAGCGGAGCTTCTGGCTTAATAG
TTGGAACACATGCTATTTTTTATGAAAGCACAGAATTTAAAAGACTAGCATATGTTATTATTGACGAACAGCATAAATTTGGAGTT
GTTCAAAGAGAAGAGCTTAAAAACAAGGGAGAAGGGGTAGACATGCTTTTAATGTCTGCAACGCCTATTCCTAGAAGTTTCGCAT
TAACACTTTTTGGTGATCTTGAAGTTTCGTTTATTAAAACTTTACCTAAGGGCCGTTTTCCTATTACTACTTATTTAGTAAGGCATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTTTTAAAAGGCCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTT
CAGAGAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTAAAAGAAGTATTTAGCGAATATGTTGTAGACATGCTTCAT
TCTAAGTTGCCATCTGATTTAAAAGAAGAAATTATGAAAAATTTTTATTCAAAAAAAGTAGATATTTTGGTAGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAGAAATTTAAATCGGGCAGAGATTCTTCTGGTAAGATTAGTATTAGAAGGAGAGGT
GGTGGGCATAAGAGAAAGTATAGGTTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTGCTTCTATTGAATA
TGATCCTAATAGAAGTGCTAATATAGCTCTGCTTGTGTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAGGGCATTAAG
GTTGGAGATGTTTTGGAAAGTGGTCCAAATGCCCCAATTAAAATTGGCAATGCTTTACCTCTTGAAAATATTCCTATTGGAAGAAC
CGTTCACAATATTGAGCTTAATATGGGAAAGGGTGGACAACTTATAAGAAGTGCTGGTGGATATGCTATGATACTTGCTTCTGAC
GGGAATTATGTCACTGTAAAATTATCATCTGGTGAGGTGAGATTGATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAA
ATGAAGATTATGCCAATATTTCTATAGGGAAAGCTGGTAAAAGCAGATGGCTTGGTAGAAGACCTAAGGTTAGAGGTGTTGCTAT
GAATCCTGTTGACCATCCACATTCAGGCAGTCTATCTGGTGGTGAGGCCCAGCGTATTAGGCTTGCTACTCAAATAGGATCAGCAC
TTTCAGGCGTTATTTATGTTCTTGATGAACCAAGCATTGGTCTTCATCAAAGAGATAATGAAAAATTGATCTCTACTCTTGTTAATC
TTAAAAATCTTGGTAATACTGTGATTGTTGTTGAACATGATGAGCAAACTTTGCGTACTGCAGATTATATTATTGATATGGGACCT
GGTGCTGGGATTCTTGGAGGAGAAATAGTTGCAAAGGGAGCCTTAATTGATATTTTAAATAGCAAGAATAGTTTGACTGGTCAAT
ATCTTAGCGGAAAGTTTAAAATAGATGTTCCAAGCTCCAGAAGAAAGACTGATAAGGGAGAAATTTTGCTTTTAGGTTCTAATAA
AAACAATCTTAAAAATATAGATGTGAGTATTCCTTTGGGAGTTTTTACCGTAATAACGGGTGTTTCTGGTAGTGGAAAAAGTACGT
TGCTTAACGAAGTGTTATATCCAGCTCTTGATAGTAGATTAAAGCTTGATGAAAAGTATTGCGATGGTTTTAAAGAT

>BcalifornensisCA443
AAAAAAGAACTTGTTATACATGATAACTTAGTATTTGATTTGATATTAAATATTAAATTATTAAAATTTAATTTGCTTGCCAATAG
AAGTACTATTGGTATATTTGCTTTTATTGGTGCTTCTGGATCAGGAAAATGCAAATTGACGGATATTCTATCACAAGAGTTTAAAA
TTCCAAAGTTTAGTCTTAACATGGGTGAGTATAGTGATTTTAATTCTCTTGATAGATTGATTGGGCCTGTTTTAAGTAATGAAGGGT
ATTATGAATCTACTAGATTTTTCAAATTTTTGAACAAATCTTCTAATTCTATTATTTTCCTATCAGATTTTGATAAATGTAATAAAA
GGGTTTTTGATTTTTTTTTAGAGGGGTTTAAAACAGGTAAACTTTTTGATGGTCTTGGGAAAAAGGTGAGCTTATCAGAAAGTTTA
ATAGTAATAAGTGTTGATGCTGATAGCAAGGAGCTTAATAGCATTGGCTTTAAAAATAAAATGGCAGGAGGAAATGATTTTAACT
TTATATTAGAGAAAAGATTTTCTAATAAATTTTTAGAGCTAATAGATCATGTGTTTGTATTTGCAGGATATGTTGGTGAAGATGTG
GAAAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGCTTGGCTGAGAAAGGGATTATTTATATAGACGAGATAG
ATAAAATTGCTAAAAAAAAATGAAAACGTTTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAACAGGCTTTGTTAAAGATAAT
TGAAGGCACCATTGCCAATGTTCCTCCAAGGGGTGGCAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATA
CTCTTTATATGTGGCGGTGCTTTTGTTGGACTTGAAAATATTGTTAAAAATCGAATAAATAAAAGTTCTATTGGGTTTTCAGCAATT
GAAAAAAAAAATATAAGAGAAGATACTTCATTAAAGTATTTAGAAATGGAAGATTTGGTTAAATTTGGCTTAATACCAGAGTTTG
TTGGTAGGCTTCCTGTGCATTCATATCTTGAAAAGTTGAATAAAGAAGATTTGTTGAGAATATTGGTTGATCCTCAAAATTCTATT
GTTAAGCAGTATTATCATATGTTTAAAATGGACAATGTTGAATTGGTATTTGAAAAAGATGCTTTAGAATCAATTGTAGATGAAGC
TTTTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACCTCTCAAAAGCCTAAAAACGTAATTTATTCTAACGTTGAATATT
ATGAAAATTACAACGCAAATGTACATAGAAGCGGGCATAAATTTGCAATTCAATCTAGCATAAAAATAGAAAAAACAAGAGAAC
TTGTAAAAAATTTCATTAATGCAGAATCTACAAAAAATATAATATTCACCTCTGGAACTACTGATGGAATTAATACTATTGCAAGC
TCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAACAGCAATTTGCTTCCTTGGGTA

AATCTTGCAAATTTAGCTAATTTAACAATTAAATTAGCTAAGTTCAATGAAATGGGAATTATTACCCCTGAAGAAATCGAAAAAC
TTATTACAGAAAAAACAAAGCTAATTAGTATTTCAGGAATAAATAATACCTTAGGAACAATTAATGATTTGGAATCTATTGGAAA
AATCGCAAAAAAATACAATATATGTCTTTTTGTAGATGCTGCACAAATGGCATATAATGCTCATGACAATTTAACAGTAATAAGC
AGCACAAAAAAAACAATAAAAGATACTATCTTAGAACAACTTGGAATAGAACATAAAAATTTTCTATCTTGTGATCTAATATTTA
CAGAATCACAACCTTCCAAAATAATAGGAACTGAAGGAGAATTTTTGGCTTCTAAAAATCTTGACAACAAATCGGGATGCCATGC
AATCATGAATTCTTATGTTCATACAAGCAATGATAAAAATAAAATAGCTGTATTTTTTGATAACGAAGAAGTAGGATCTTTAACCT
CAAGAGGTGCTGATTCAAATTTTTTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACCAGAGAAGAGCATTTAATA
AAAACAAACAAATCATTTAATATCTCGATTGACAGCGTTCACGGCATTCATCCAGGATATGCATCTAAACATGATCCAAACTATC
AAGCAACTCTGGGTAAAGGCGTATTTGTAAAAAATAGCGCCAATTTCAGATATGCAACAACTTCAACAGGATTTGCAAAATTAAA
AAATTTAGCTATTAAAAATAATATTGGAAGTGGCAATATTGCTTTTATTCATTTAACCTATGTACCAAGTCCAGCTGGAATTAATG
AGCAAAAGTCTAAACCTACTCAACAGAGTGTTAAAACCCTAAATAAAGCGGGCATTTTTCCTGATTTAATTATTGCTAGAAGCTCT
CAAGTATTGACAGATCAAATCAGAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGTACTTCTATTATTGACAATGTTGATGTTTC
TACTATTTATGAAATTCCTATATCTTTTTACAAACAAGGGGTACATGAAATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCCAATTTTTTTGCGCCTAAAAAAATTATTAATATTGCTGT
TTGTGGTAAATATGCTGAACTTGATGATTCTTATGCATCAATTAGAGAGTCCTTGGTTCATGTTGCAGCTAATTTAGATTTGCTTAT
TAAAAGCACTTTAATTGATTCTAATGATTTGAATGAGAGCTGTTTAAAGGAGTTTGACGGTATTATTGTTCCTGGCGGCTTTGGGG
GCAAAGGATATGAGGGTAAAATTATGGCTATTGCTCGTCAACATTATGATAATTTATCCAACATATTGTCGTCTTTTAACATTTCA
GTAACTCTTTTGACTGGTAGTTTAAAAAAGAAGGATAAGGAGCAAGCATTAGAAAGTATTAGAAATGGAACTTCCGGTTTAATAG
TTGGAACACATGCTATTTTTTATGAAAGCACAGAATTTAAAAGATTAGCATATGTTATCATTGACGAGCAGCATAAGTTTGGAGTT
ATCCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGTGTTGATATACTTTTAATGTCTGCAACGCCTATTCCTAGAAGTTTTGCGT
TAACACTTTTTGGTGATCTTGAAATTTCGTTTATTAAAACCTTGCCTGAGGGTCGTTTGCCCATTACGACTTATTTAGCAAAGCATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTTTTAAAAGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCGTCTT
CAGAGAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTGAAGGAAGTATTTGGCGAATATGTCGTAGATATGCTTCAT
TCTAAGTTGCCATCCGATTTAAAAGAAGAAATTATGAAAAATTTTTATTTAAAAAAAGTAGATATTTTGGTGGCTACTAGTGGTAA
CGATCCTTTGAAATCTTTAACAAAAGGTAAGAAATTTAGATCGGGCAGAGATTCTTCTGGCAGAATTAGTGTTAGAAGAAGAGGT
GGTGGACATAAGAGAAGGTATAGGTTGATCGATTTTAATCGAAGAGATAAATTTAGCGTTCCTGCTCGAGTTTCTTCTATTGAATA
TGATCCCAATAGAAGTGCTAATATAGCTTTGCTTATTTATAAAGATGGAGAAAAAAGGTATATTATTTCTCCTAAAGGCATTAAGG
TTGGAGATGTTTTGGAAAGTGGTCCAAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAACT
GTTCACAATATTGAGCTTAATGTGGGAAAAGGTGGACAGCTTATAAGAAGTGCTGGTGGATATGCTATGATACTTGCTTCTGACG
GGAATTATGTTACTGTAAAATTATCATCCGGTGAGGTGAGATTGATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAAAT
GAAGATTATGCCAATATTTCTATAGGTAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGGCCTAAGGTTAGAGGCGTTGCTATGA
ATCCTGTTGACCATCCGCATTCAGGCAGCCTATCTGGTGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGCTCAGCACTT
TCAGGCGTTATTTATGTTCTTGATGAGCCAAGTATTGGTCTTCATCAAAGAGATAATGAAAAATTAATCTCTACTCTTGTTAATCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAGCATGATGAGCAAACTTTACGTACTGCAGACTATATTATTGATATGGGGCCTGG
TGCTGGAATTCTTGGAGGGGAAATAGTTGCAAAGGGAGCCTTAATTGATATTTTAAATAGCAAGAATAGCTTAACTGGTCAATAT
CTTAGCGGTAAGTTTAAAATAGATGTTCCAAGCTCTAGAAGAAAGGCAGATAAGGGAGAAATTTACTTTTAGGTTCTAATAAAA
ACAATCTTAAAAATATAGATGTAAGTATCCCTTTGGGAGTTTTTACCGTAATAACAGGTGTTTCTGGTAGTGGAAAAAGCACTTTA
CTTAACGAAGTGTTATACCCAGCTCTTGATAGTAGATTAAAGTTTAATGAAAAGTATTGTGATGGTTTTAAAGAC

>Strain25015
AAAAAAGAACTTATTATGCATGATAACTTGATATTTGATTTGATATTAAATATTAAATTATTAAAATTCAATTTGCTTGCCAACAG
AAGTACTATTGGTATATTTGCCTTTATTGGTGCTTCTGGATCAGGAAAATGCAAGCTGACAGATATTTTGTCACAAGAGTTTAGAA
TTCCAAAGTTTAGTCTTAACATGGGTGAGTATAGTGATTTTAATTCTCTTGATAGATTGATTGGGCCTGTTTTAAGCAATGAAGGG
TATTATGAATCTACTAGATTTTTTAAATTTTTAAACAAATCTTCTAATTCTATTATTTTCCTATCAGATTTTGACAAATGTGATAAA
AGGGTTTTTGATTTTTTTTTGGAAGGGTTTAAAACAGGCAAGCTTTTTGATGGTCTTGGGAAAAGTGTGAGTTTATCAGAAAGTTT
AATAGTAATAAGTGTCAATGCTGAAAGTAAGGAGCTTAATAGTATTGGCTTTAAAAATAAAATGGTGGGGGAAAATGATTTTAAT
TTTATATTAGAGAAGAGATTTTCTAATAAGTTTTTAGAGTTAATAGATCATGTGTTTGTATTTGCAGGATATGTTGGTGAAGATGT
AGAAAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGTTTAGCTGAGAAAGGTATTATTTATATAGACGAGATAG
ATAAAATTGCTAAAAAAAAATGAAAATGTTTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAACAGGCTTTGTTAAAGATAAT
TGAAGGTACCATTGCCAATGTTCCCCCAAGAGGCGGCAGAAAACATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATA
CTTTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAATATTGTTAAGAATAGAATAAATAAAAGTTCTATTGGGTTTTCAGCAATT
GAAAAAAAGAATATAAGAGAAGATACTTCATTAAAATATTTAGAAATGGAAGATTTGATTAAATTTGGCTTAATACCAGAGTTTG
TTGGCAGGCTTCCTGTGCATTCGTATCTTGAAAAGTTAAATAAAGAAGATTTGTTGAGAATATTGGTTGATCCCCAAAATTCTATT
GTTAAGCAGTATTATCATATGTTTAGAATGGACAATGTTGAACTGGTATTTGAAAAAGACGCTTTAGAATCAATTGTAGATGAAG
CCTTTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACTTCTCAAAAGCCCAAAAACGTAATTTATTCTAACGTTGAATAT
TATGAAAATTACAACGCAAATGTGCACAGAAGCGGGCACAAATTTGCAATTCAATCGAGCATAAAGATAGAAAAAACAAGAGAA
CTTGTAAAAAATTTCATTAATGCAGAATCTGCAAAAAATATAATATTCACCTCTGGAACTACTGACGGAATCAATACTATTGCAAG
CTCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAACAGTAATTTGCTTCCTTGGGT
AAATCTTGCAAACTTGGCTAATTTAACAATTAAATTAGCCAAGTTCAATGAAATGGGCATTATTACCCCTGAAGAAATTGAAAAA
CTTATTACAGAAAAAACAAAGCTAATTAGTATTTCAGGAATAAATAATACCTTAGGAACCATTAATGATTTGGAATCTATTGGAA
AAATCGCAAAAAAATACAATATATGTCTTTTTGTAGATGCTGCACAAATGGCATATAATACTCATGACAATTTAACAGTAATAAG
TAGCACAAAAAAACCACTAAAAAATACTATCTTAGAACAACTTGGAATAGAACATAAAGATTTTCTATCTTGTGATCTAATATTTA
CAGAATCACAACCTTCTAAAATAATAGGAACTGAAGGAGAATTTTTGGCTTCTAAAAATCTTGACAACAAATCGGGATGCCATGC
GATCATGAATTCTTATGTTCATACAAGCAATGATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAAGTAGGATCTTTAACCT
CAAGAGGCGCTGATTCAAATTTTTTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACCAGAGAAGAGCATTTAATA
AAAACAAACAAATCATTTAATATTTCAATTGATAGCGTTAACGGCATTCATCCGGGATATACATCCAAACATGATCCAAACTATC
AAGCAACTCTAGGAAGGGGTATGGTTGTAAAAAATAGTGCCAATTTCAGATATGCAACAACTTCAACGGGATTTGCAAAATTAAA
AAATTTGGCTATTAAAAATAATATTGGAAGTGACAATATTTCTTTTATTCATTTAACCTATGTGCCAAGCCCGGCCGGAATTAATG
AACAAAAATCTAAGCCTACTCAACAGAGTGTTAAAACCTTAAATAAAGCAGGTATTTTCCCCGATTTAATTATTGCTAGAAGCTCT
CAAGTATTGACAGATCAAATCAGAAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGTACTTCTATTATTGACAATGTTGATGTTTC
TACTATTTATGAAATTCCTATATCTTTTTACAAACAAGGTGTACATAAGATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCTAATTTTTTTGCGCCTAAAAAAATCATTAATATTGCTGT
TTGTGGTAAATATGCTGGACTTGATGATTCTTATGCATCAATTAGAGAGTCCTTGGTTCATGTTGCAGCTAATTTAGATTTGCTTAT
TGAAAGTACCATAATTGATTCTAATGATTTGAATGAAAACCGTTTAAAGGAATTTGACGGCATTATTGTTCCTGGCGGCTTTGGAG
GTAAAGGATATGAGGGTAAAATTATGGCCATTGCTCGTCAACATTATGATAATTTATCCAACATATTGTCGTCTTTTAACATTTCA
GTCACTCTTTTGACTGGTAGTTTAAAAAAGAAGGATAAGGAACAGGCATTGGAAAATATTAGAAATGGAACTTCCGGCTTAATAG
TTGGGACACATGCTATTTTTTATGAAAGCACAGAATTTAAAAGATTGGCATATGTTATCATTGACGAGCAGCATAAGTTTGGAGTT
GTCCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGTGTAGACATGCTTTTAATGTCTGCAACGCCTATTCCTAGGAGTTTTGCGT
TAACGCTTTTTGGTGATCTTGAAATTTCTTTTATTAAAACCTTGCCTGAGGGCCGTTTGCCTATTACTACTTATTTAGCAAAGCATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAGGGCCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTT
CAGAGAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTGAAGGAAGTATTTGGCGAATATGTTGTAGACATGCTTCAT
TCTAAGTTGCCATCTGATTTAAAAGAAGAAATTATGAAAAATTTTTATTTAAAAAAAGTAGATATTTTAGTGGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCGGGCAGAGATTCTTCTGGTAGGATTAGTGTTCGAAGAAGAGGT
GGTGGGCATAAGAGAAGGTATAGGTTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTTCTTCTATTGAGTA
TGATCCTAATAGAAGCGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGATATATTATTTCTCCTAAAGGCATTAAGG
TTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAACT

GTTCACAATATTGAGCTTAATGTAGGAAAGGGTGGGCAGCTTGTAAGAAGTGCTGGTGGATATGCTATGATACTTGCTTCTGACG
GGAATTATGTCACTGTAAAATTATCATCTGGTGAGATGAGATTGATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAAAT
GAAGATTATGCCAATATTTCTATAGGGAAAGCTGGCAAAAGTAGGTGGCTTGGTAGAAGGCCCAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGACCATCCACATTCAGGCAGCCTATCTGGTGGTGAGGCTCAGCGCATTAGGCTTGCTACTCAAATAGGCTCAGCACTT
TCAGGCGTTATTTATGTTCTTGATGAGCCAAGTATTGGCCTTCATCAAAGAGATAATGAAAAATTAATCTCTACTCTTGTTAATCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAGCATGATGAGCAAACTTTGCGCACCGCAGACTATATTATTGATATGGGTCCTGG
TGCTGGAATTCTTGGAGGGGAAATAGTTGCAAAGGGAGCCCTAATTGATATTTTAAATAGCAAAAACAGCTTAACTGGCCAATAT
CTTAGTGGTAAGTTTAAAATAGATATTCCAAGCTCTAGAAGAAAGACAGATAAGGGAGAAATTTTACTTTTAGGTTCTAATAAAA
ATAATCTTAAAAATATAGATGTAAGTATCCCTTTGGGAGTTTTTACCGTAATAACAGGTGTTTCTGGTAGCGGAAAAAGTACTTTA
CTTAACGAAGTATTGTATCCAGCTCTTGATAGTAGATTAAAGCTTAATGAAAAGTATTGTGATGGTTTTAAAGAC

>BcarolinensisscW22
AAAAAAGAACTTATTATACATGATAACTTGGTATTTGATTTAATATTAAATATTAAATTATTAAAATTCAATTTGCTTGCCAACAG
AAGTACTATTGGTATATTTGCCTTTATTGGTGCTTCTGGATCAGGAAAATGCAAATTGACGGATATTTTGTCACAAGAGTTTAAAA
TTCCAAAGTTTAGTCTTAACATGGGTGAGTATAGTGATTTTAATTCTCTTGATAGATTGATTGGGCCTGTTTTAAGTAATGAAGGGT
ATTGTGAATCTACTAGGTTTTTCAAATTTTTAAACGAATCTTCTAATTCTATTATTTTCTTATCAGATTTTGATAAATGTGATAAAA
GGGTTTTTGATTTTTTTTTGGAGGGGTTTAAAACAGGAAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCAGAAAGTTTA
ATAGTAATAAGCGTCAATGCTGAGAGCAAGGAGCTTAATAGTATTGGCTTTAAAAATAAAATGATGGGGGAAAATGATTTTAACT
TTATATTAGAGAAGAGATTTTCTAATAAGTTTTTAGAGTTAATAGATCATGTGTTTGTATTTGCCGGATATGTTGGTGAAGATGTA
GAAAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGCTTAGCTGAGAAAGGGATTATTTATATAGACGAGATAG
ATAAAATTGCTAAAAAAAAATGAAAATGTCTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAACAGGCTTTGTTAAAGATAAT
TGAAGGCACCGTTGCCAATGTTCCTCCAAGAGGCGGTAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATA
CTCTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAATATTGTTAAGAACCGAATAAATAAAAGTTCTATTGGGTTTTCAGCAATT
GAAAAAAAGAATATAAGAGAAGATACTTCATTAAAGTATTTGGAAATGGAAGATTTGATTAAATTTGGCTTAATACCAGAGTTTG
TTGGCAGACTTCCTGTGCATTCGTATCTTGAAAAGTTAAATAAAGAAGATTTGTTGAGAATATTGGTTGATCCCCAAAATTCTATT
GTTAAGCAGTATTATCATATGTTTAAAATGGATAATGTTGAATTGGTATTTGAAAAGGATGCTTTAGAATCAATTGTAGATGAAGC
CTTTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACTTCTCAAAAGCCCAAAAACGTAATTTATTCTAACGTTGAATATT
ATGAAAATTACAACGCAAATGTGCACAGAAGCGGGCACAAATTTGCAATTCAATCGAGCATAAAGATAGAAAAAACAAGAGAAC
TTGTGAAAAATTTCATTAATGCAGAATCTGCAAAAAATATAATATTCACCTCTGGAACTACTGATGGAATCAATACTATTGCAAGC
TCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAACATAACAGTAATTTGCTTCCTTGGGTA
AATCTTGCAAGTTTGGCTAATTTAACAATTAAATTAGCCAAGTTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAAC
TTATTACAGAAAAAACAAAACTAATTAGTATTTCAGGAATAAATAATACCTTGGGAACCATTAATGATTTGGAATCTATTGGAAA
AATAGCAAAAAAATACAATATATGTCTTTTTATAGATGCTGCACAAATGGCATATAATACTCATGACAATTTAACAGTAATAAGC
AGCACAAAAAAACCACTAAAAAATGCTATCTTAGAACAACTCGAAATAGAACATAAAAATTTTCTATCTTGCGATCTAATATTTA
CAGAATCACAACCTTCTAAAGTAATAGGAACTGAAGGGGAATTTTTGGCTTCTAAAAATCTTGATAACAAATCGGGATGCCATGC
GATCATGAATTCTTATGTTCATACAAGCAATGATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAGATAGGATCTTTAACCT
CAAGAGGTGCTGATTCAAATTTTTTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACCAGAGAAGAGCATTTAATA
AAAACAAACAAATCATTTAATATCTCAATTGACAGCGTTCACGGCATTCATCCGGGATATACATCCAAACATGATCTAAACTACC
AAGCAACTCTGGGAAGGGGCATGGTTGTAAAAAATAGCGCCAATTTCAGATATGCAACAACTTCAACAGGATTTGCAAAATTAAA
AAATTTGGCTATTGAAAATAATATTGGAAGTGGCAATATTTCTTTTATTCACTTAACCTATGTGCCAAGCCCGGCCGGAATTAATG
AACAAAAATCTAAGCCTACTCAACAGAGTGTTAAAACCTTAAATAAAGCAGGTATTTTCCCTGATTTAATTTGCCAGAAGCTCT
CAAGTACTAACAGATCAAATCAGAAAAAAAGTGGCAATGTTTTGCAATGTTGAGAGCACTTCTATTATTGACAATGTTGATGTTTC
TACTATTTATGAAATTCCTATATCTTTTTACAGACAAGGTGTACATAAGATTTTAAGCTCTAAGTTAAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAAGCTTGTAGGAGTTATAAAATCTAATTTTTTTGCACCTAAAAAAATTATTAATATCGCTGT
TTGTGGTAAATATGCTGAACTTGATGATTCTTATGCATCAATTAGAGAGTCCTTGGTTCATGTTGCAGCTAATTTAGATTTGCTTAT
TAAAAGCACCATAATTGATTCTAATGATTTAAATGAAAACCTTCTAAAGGAGTTTGACGGCATTATTGTTCCTGGTGGCTTTGGAG
GCAAAGGATATGAGGGGAAAATTATGGCTATTGCTCGTCAACATTATGATAATTTATCCAATATATTGTCGTCTTTTAACATTTCA
GTGACTCTTTTGACTGGTAGTTTAAAAAAGAAAGATAAGGAACAAGCATTAGAAAATATTAGAAATGGAACTTCCTGCTTAATAG
TTGGAACACATGCTATTTTTTATGAAAGTACAGAATTTAAAAGATTAGCATATGTTATCATTGACGAGCAGCATAAGTTTGGAGTT
GTTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGTGTAGATATGCTTTTAATGTCTGCAACGCCTATTCCTAGGAGTTTTGCGT
TAACGCTTTTTGGTGATCTTGAAATTTCTTTTATTAAAACCTTGCCTGAAGGTCGTTTGCCTATTACTACTTATTTAGCAAAGCATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAAGGTCATCAGGTTTATTTTATTTATCCATTAATTTCATCTT
CAGAGAAATTTGAATTAAAAGATGTTAATAATATGTTTAAAATTGAAGGAAGTATTTGGCGAATATGTTGTAGACATGCTTCAT
TCTAAGTTGTCATCCGATTTAAAAGAAGAAATTATGAAAAATTTTTATTTAAAAAAAGTAGATATTTTGGTAGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCGGGCAGAGATTCTTCTGGTAGGATTAGTGTTAGAAGAAGGGGT
GGTGGGCATAAGAGAAGGTATAGGTTGATTGATTTTAATCGGAGAGATAAATTTAGCATTCCTGCTCGAGTTTCGTCTATTGAATA
TGATCCTAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGATATATTATTTCTCCTAAAGGCATTAAGG
TTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAACT
GTTCACAATATTGAGCTTAATGTGGGAAAGGGTGGGCAGCTTGTAAGAAGTGCTGGTGGATATGCTATGATACTTGCTTCTGAAG
GGAATTATGTCACTGTAAAATTATCATCTGGTGAGATGAGATTGATTTTTAAAAAATGTATTGCAACAATTGGTGAAATTGGAAAT
GAAGATTATGCCAATATTTCTATAGGGAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGGCCCAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGACCATCCACATTCAGGCAGCCTATCTGGTGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGCTCAGCACTT
TCAGGCGTTATTTATGTTCTTGATGAGCCAAGCATTGGCCTTCATCAAAGAGATAATGAAAAATTAATCTCTACTCTTGTTAATCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAGCATGATGAGCAAACTTTACGCACCGCAGACTATATTATTGATATGGGGCCTGG
TGCTGGAATTCTTGGAGGGGAAATAGTTGCAAAGGGAGCCTTAATTGATATTTTAAATAGCAGAAATAGCTTAACTGGTCAATAT
CTTAGTGGTAAGTTTAAAATAGATGTTCCAAGCGCTAGAAGAAAGACAGATAAGGGGGAAATTTTACTTTTAGGTTCTAATAAAA
ATAATCTTAAAAATATAGATGTAAGTATTCCTTTGGGAGTTTTTACCGTAATAACAGGTGTTTCTGGTAGCGGGAAAAGCACTTTA
CTTAACGAAGTTTTGTATCCAGCTCTTGATAGTAGATTAAAGCTTAATGAAAAGTATTGTGATGGTTTTAAAGAC

>BbissettiiDN127
AAAAAGGAACTTATTATACATGATAACTTAGTATTTGATTTAATATTAAATATTAAATTATTAAAATTCAATTTGCTTGCCAACAG
AAGTACTATTGGTATATTTGCCTTTATTGGTGCTTCTGGATCAGGAAAATGCAAATTGACGAATATTTTGTCACAAGAGTTTAAAA
TTCCAAAGTTTAGTCTTAACATGGGTGAATATAGTGATTTTAATTCTCTTGATAGATTGATTGGGCCTGTTTTAAGTAATGAAGGGT
ATTATGAATCTACTAGGTTTTTCAAATTTTTAAACAAATCTTCTAATTCTATTATTTTCTTATCAGATTTTGATAAATGCGATAAAA
GGGTTTTTGATTTTTTTTTGGAGGGGTTTAAAACAGGCAAACTTTTTGATGGTCTTGGGAAAAAGGTAAGTTTATCAGAAAGTTTA
ATAGTAATAAGCCTCAATGTTGAGAGCAAGGAGCTTAATAGTATTGGCTTTAAAAATAAAATGATGGGGGAAAATGATTTTAACT
TTATATTAGAGAAGAGATTTTCTAATAAGTTTTTAGAGTTAATAGATCATGTGTTTGTATTTGCCGGATATGTTGGTGAAGATGTA
GAAAATATTTTGCTTAAATTGATACATGCTGCTCATGGGGATGTTAGCTTAGCTGAGAAAGGGATTATTTATATAGACGAGATAG
ATAAAATTGCTAAAAAAAAATGAAAATGTCTCAATAACAAGAGATGTTTCTGGAGAAGGGGTTCAACAGGCTTTGTTAAAGATAAT
TGAAGGCACCATTGCCAATGTTCCCCCAAGAGGTGGTAGAAAGCATCCTTATGAGGATACTATTGAAATTAATACTCAAAATATA
CTCTTTATATGTGGTGGTGCTTTTGTTGGGCTTGAAAATATTGTTAAGAACCGAATAAATAAAAGTTCTATTGGGTTTTCAGCAATT
GAAAAAAAGAATATAAGAGAAGATACTTCATTAAAGTATTTGGAAATGGAAGATTTGATTAAATTTGGCTTAATACCAGAGTTTG
TTGGCAGACTTCCTGTGCATTCGTATCTTGAAAAGTTAAATAAAGAAGATTTGTTGAGAATATTGGTTGATCCCCAAAATTCTATT
GTTAAGCAGTATTATCATATGTTTAAAATGGACAATGTTGAATTGGTATTTGAAAAAGATGCTTTAGAATCAATTGTAGATGAAGC
CTTTGACAATAAGCATATAATTTATTTTGACAATGCAGCAACTTCTCAAAAGCCCAAAAACGTAATTTATTCTAACGTTGAATATT
ATGAAAATTACAACGCAAATGTGCACAGAAGCGGGCACAAATTTGCAATTCAATCGAGCATAAAGATAGAAAAAACAAGAGAAC

TTGTGAAAAATTTCATTAATGCAGAATCTGCAAAAAATATAATATTCACCTCTGGAACTACTGATGGAATCAATACTATTGCAAGC
TCATTTTTTTACTCAAAATACTTTAAAAAAAAAGATGAAATTATTCTTACAACTCTTGAGCATAACAGTAATTTGCTTCCTTGGGTA
AATCTTGCAAGTTTGGCTAATTTAACAATTAAATTAGCCAAGTTCAATGAAATGGGAATTATTACCCCTGAAGAAATTGAAAAAC
TTATTACAGAAAAAACAAAACTAATTAGTATTTCAGGAATAAATAATACCTTAGGAACCATTAATGATTTGGAATCTATTGGAAA
AATCGCAAAAAAATACAATATATGTCTTTTTATAGATGCTGCACAAATGGCATATAATGCTCATGACAATTTAACAATAATAAGC
AGCACAAAAAAACCAATAAAAGATAATATATTAGAACAACTTGGAATAGAACATAAAAATTTTCTATCTTGTGATCTAATATTTA
CAGAATCACAACCTTCCAAAATAATAGGAACTGAAGGAGAATTTTTGGCTTCTAAAAACCTTGACAACAAATCGGGATGCCATGC
AATCATGAATTCTTATGTTCATACAAACAATGATAAAAATAAAATAGCTGTATTTTTTGACAACGAAGAGATAGGATCTTTAACCT
CAAGAGGCGCTGATTCAAATTTTTTATCAGAAGTTTTAGAAAGAATCGATCTTGCTCTTAATTTAACCAGAGAAGAGCATTTAATA
AAAACAAGCAAATCATTTAACATCTCAATTGACAGCGTTCACGGCATTCATCCGGGATATACATCCAAACATGATCCAAACTACC
AAGCAACTCTGGGAAGGGGCATGGTTGTAAAAAATAGTGCCAATTTCAGATATGCAACAACTTCAACAGGATTTGCAAAATTAAA
AAATTTGGCTATTAAAAATAATATTGGAAGTGGTAATATTGCTTTTATTCACTTAACCTATGTGCCAAGCCCGGCCGGAATTAATG
AACAAAAGTCTAAACCTACTCAACAGAGTGTTAAAACTTTAAATAAAGCAGGCATTTTCCCTGATTTAATTATTGCTAGAAGTTCT
CAGGTATTGACAGATCAAATCAGAAAAAAAGTGGCAATGTTTTGTAATGTTGAGAGCACTTCTATTATTGATAATGTTGATGTTTC
TACTATTTATGAAATTCCTATATCTTTTTACAAACAAGGTGTACATGAGATTTTAAGTTCTAAGTTGAATATTAAGGTT---
GATCCAAAAATAGAAGAGCTTTCAAACCTTGTGGGAGTTATAAAATCTAATTTTTTTGCGCCTAAAAAAATTATTAATATTGCTGT
TTGTGGTAAATATGCCGAACTTGATGATTCTTATGCGTCAATTAGAGAGTCTTTGGTTCATGTTGCAGCTCATTTGGACTTGCTTAT
TAAAAGTACTTTAATTGATTCTAATGATTTAAATGAGAACTGTTTAAAAGAGTTTGATGGCATTATTGTTCCTGGTGGCTTTGGGG
GAAAAGGATATGAAGGTAAAATTATGGCTATTGCTCGTCAACATTATGATAATTTATCCAACATATTGTCGTCTTTTAACATTTCA
GTAACTCTTTTGACTGGTAGTTTAAAGAAGAAGGATAAGGAACAAGCATTAGAAAATATTAGAAATGGAACTCCCGACTTAATAG
TTGGAACACATGCTATTTTTTATGAAAGCACGGAATTTAAAAGATTAGCATATGTTATCATTGACGAGCAGCATAAGTTTGGAGTT
GTTCAAAGAGAAGAGCTTAAAAACAAAGGAGAAGGTGTAGATATGCTTTTAATGTCTGCAACGCCCATTCCTAGAAGTTTTGCGT
TAACGCTTTTTGGTGATCTTGAAATTTCTTTTATTAAAACCTTGCCTGAAGGCCGTTTGCCTATTACTACTTATTTAGCAAAGCATG
GCAATGAAGATAAAGTTTATGAGTTTTTAAGAAAAGAGCTCTTAAAGGGTCATCAGGTTTATTTTGTTTATCCATTAATTTCATCTT
CAGAGAAATTTGAATTAAAAGATGTTAATAATATGTGTTTAAAATTGAAGGAAGTATTTGGCGAATATGTTGTAGACATGCTTCAT
TCTAAGTTGTCGTCCGATTTAAAAGAAGAAATTATGAAAAATTTTTATTTTAAAAAAGTAGATATTTTGGTGGCTACTAGTGGTAA
TGATCCTTTGAAATCTTTAACAAAAGGTAAAAAATTTAAATCGGGCAGAGATTCTTCTGGTAGGATTAGTGTTAGAAGAAGGGGT
GGTGGGCATAAGAGAAGGTATAGGTTGATTGATTTTAATCGAAGAGATAAATTTAGCATTCCTGCTCGAGTTTCGTCTATTGAATA
TGATCCTAATAGAAGTGCTAATATAGCTTTGCTTGTTTATAAAGATGGAGAAAAAAGATACATTATTTCTCCTAAAGGCATTAAGG
TTGGAGATGTTTTGGAAAGTGGTCCGAATGCCCCAATTAAAATTGGCAATGCCTTGCCTCTTGAAAATATTCCTATTGGAAGAACT
GTTCACAATATTGAGCTTAATGTGGGAAGGGGTGGACAGCTTGTAAGAAGTGCTGGTGGATATGCTATGATACTTGCTTCTGAAG
GGAATTATGTCACTGTAAAATTATCATCTGGTGAGATGAGATTGATTTTTAAAAAATGTATTGCAACAATTGGTGAGATTGGAAAT
GAAGATTATGCCAATATTTCTATAGGTAAAGCTGGTAAAAGTAGGTGGCTTGGTAGAAGGCCCAAGGTTAGAGGTGTTGCTATGA
ATCCTGTTGATCACCCACATTCAGGCAGCCTATCTGGTGGTGAGGCTCAGCGTATTAGGCTTGCTACTCAAATAGGCTCAGCACTT
TCAGGCGTTATTTATGTTCTTGATGAGCCAAGCATTGGTCTTCATCAAAGAGATAATGAAAAATTAATCTCTACTCTTGTTAATCTT
AAAAATCTTGGCAATACTGTAATTGTTGTTGAGCATGATGAGCAAACTTTGCGCACCGCAGACTATATTATTGATATGGGTCCTGG
TGCTGGAATTCTTGGAGGAGAAATAGTTGCAAAGGGAGCCCTAATTGATATTTTAAATAGCAAGAATAGTTTAACTGGTCAATAT
CTTAGTGGTAAGTTTAAAATAGATGTTCCAAGCTCTAGAAGAAAGGCAGATAAGGGGGAAATTTTACTTTTAGGTTCTAATAAAA
ATAATCTTAAAAATATAGATGTAAGTATTCCTTTGGGAGTTTTTACCGTAATAACAGGTGTTTCTGGTAGCGGGAAAAGCACTTTG
CTTAACGAAGTGTTATATCCAGCTCTTGATAGTAGATTAAAGCTTAATGAAAGGTATTGTGATGGTTTTAAAGAC

>Bduttonii
AGAACAAGTATGATAATTGATGAGAGTATTTTATCCGAATTACTTTTGCATATTAAGCTTTTAAGAATTAAATTTTTCTTTAAAAAT
AGTACTCTTGGTATTTTCATTTTAATTGGATCTTCTGATGTAGATAAGAATAAACTTGCGTGTGTTTTATCAGAAGAACTTAATATG
CCTAAACTTACTTTGGGAATGAGTGAATATTGTGATTTTGATGGTATTAATCGATTGATAGGACCAGTATATGGTTCTGAAGCTTA
TGATGAACCTACTAAATTCTTTAGATTTTTAAATGAGTCTTTAAATTCAATTATTGTTTTGTCAGATTTTGATAAGTCTTCTAAGAG
AGTTATAGATTTTTTTTTTTGAAGGATTTAATACAGGTAGACTTTTTGTAATTTTGGTAGAAGTGTCAGTTTGTCAGGTAGCATTAT
AATTATAGATATTAATATAGAATATAGAGAACTTGATGGTATTGGATTTAAAAATGAAACTGTT---------
GATATCAGAAGTTTATTAGAAAAACGTTTTTCTAATAAATTTCTAGACTTAGTAGATCATGTTTTTTTCTTTGCTGGATATGTGGGA
GAGGATGTTGAGAATATTTTGCTTAAACTAATTCATGCTGCTAATGGAGATGTGAGTTTTGCAGAAAGAGGTATCATTTATATAGA
TGAGATAGATAAGATTGCAAAAAAAGGTGAGAATGTTTCAATTACAAGAGATGTATCTGGAGAGGGTGTTCAACAGTCTTTACTT
AAGATAATTGAAGGTACTATTGCAAATGTTCCACCAAGAGGTGGTAGAAAACATCCTTATGAAGAAACTATTGCAATTAATACTC
ACGATATATTATTTATATGTGGTGGAGCTTTTGTTGGACTTGAAAATATTATTAAGAAAAGAATTAATAGGAGTTTTATTGGTTTTT
CATCTTCTAGTTGTAAGGATACAGGTGGAGATAATTCTTTAAAATATTTAGAAATGGAAGATTTGATTAAATTTGGTCTTATTCCA
GAGTTTGTAGGTAGACTTCCTGTGCATTCTTATCTTGATAAATTAGAAAAAAAAGATTTGATGAAAATATTAGTTGAGCCCGAGAA
TTCTATTGTAAGGCAATACTATCATATGTTTAAGATGGATAATGTTGATCTTTTATTTGAAAAAGATGCTCTTGATGCAATTGCAG
AAGAAGCTATTAATAATAAAAAAATCATTTATTTTGACAATGCCGCGACTTCTCACAAACCTCAAAGTGTAATTTCATCTATAGTT
GAATACTACACAAATTATAATGCAAATGTTCATAGAAGTGGACATCAATTTGCAATAGAATCAAGCTTAAAAATAGAAGGAGACA
AGAAAAATTGTTAAACAATTTATCAATGCAGAATCTCCTAAAAACATAATTTTCAATTCCGGAACTACAGATGGAACAAATACAA
TAGCAAATTCATTATTGTTATCAAAATTTTTTAAAGAAAATGATGAAATTATTACAACAACACTTGAACATAATAGTAATTTACTT
CCTTGGATCAACATTGCAAAATTTTTAAACTTAAAAATTAAACTTGCAAAATTCAATGAAATGGGAATTATTCAATCAGAACAAA
TACAAAATCTAATTACAGATAAAACAAAAATCATTACTATGTCTGGAATAAATAATATACTAGGAACAATGCAAAATCTAGAGCA
AATAGGAAAAATTGCCAGCGATAACAAAATTACATTATTCATAGATGCGGCTCAAATGGCATATAACACTCATGAAAATATAGTT
GTTATAACAAGTTTTAAAAAAAGTATTAAAGAAATTATCCTAGAAAAATTACAAATATCTCAACAAGATTTCTTATCATGTGATTT
AATATTCACAGCCTCTGAACCTTCTAAAATTATAGGCAGTGAAGGTGAATTTTTAGCATCAAAAAATCTTGACAACAAATCAGGA
TGTCATGCCATAATGAATGCATTTGTTCACACAAATAACAAAAAAAACAAAGTGGCTGTATTTTTTGACAATGAAGAAATTGGAT
CTCTAACGTCTAGAGGAGCTAATTCAACACTACTAACAGAAATTTTAGAAAGAATAGATTATGTCTTAAATTTAGGACAAGAAGA
ACATATGATCAAACTCCAAAAATCATTTAATATCTCAATGGATGGAGCACATGGAATACATCCAGGATACACATGTAAACATGAC
CCCTACTACAAAACAAGTCTAGGGAAAGGTGTGGCTATTAAAAGCAATGCCAATTTTAAATATGCAACAACAGCAAATGGATGG
GCAAAACTTAAAGCTTTAGCTATAAAAAATAATATTGGAATTGATAATATTGCTTTTGTTCATTTAACTTATATTCCAAATCCTGCT
GGAATAAATGAACAAAAATCTAAACCTACTCAACAGAGTGTTAAGACCTTAAATAAAGCTGGAATTTTTCCTGATTTAATTATTGC
AAGAAGTTCTCAACTCTTAACAAGGCAAATAAGAGATAAGATAGCAATGTTTTGTAATGTTGATGCCTCTTGCATTGTTGATAATA
TTGATGTTGCAACTATTTATGAAATACCTATATCTTTTTATAAACAAGGATTGCATAAAATTTTAGGTTCAAAACTGAAAATTAAT
GTT---
ACACCTAAGGTGGATGGTCTTGAAAAGTTAGTAAATGTAATAAAGAAAAATCTTCTCTCGCCTCAAAAGGTTGTAAATATTGCTA
TTTGTGGTAAATATACTGAACTTGGTGATTCTTATGCATCAATATTTGAGTCTTTAACTCATGTTTCTGCCAATTTGGATATCTTAG
TTAAGACAACTGTCATTGACGGTATTAATTTTGATGAGAAAATGATGGAAGGTGTAGATGGCATAGTTGTGCCTGGGGGATTTGG
AAGCCGAGGGTATGAAGGTAAGATTCGTGCAATTGCACGTCAACATTATAATAGTTTGACAAATATATTAAAAGATTTTGATATTT
CTGTGGTTCTTTTGACTGGAAGTTTGAAAAAAAAAGATAGAGATGAGGTTTTAGAAAAGATTCAAAGTGGAATTTGTAGTTTGGT
AATTGGGACTCATGCTATTTTTTCTCAAGGGACAAAGTTTAAAAATTTGGCTTATGTCATTATTGATGAACAGCATAAGTTTGGAG
TTGAACAAAGAGAAGAGCTTAGAAATAAAGGTGAAGAAGTTGATGTTCTTTTAATGTCAGCAACGCCTATTCCTAGAAGCTTAGC
TTTAACCCTTTTTGGAGATCTTGAAGTGTCTTTAATTAAAAGGGGTCCTGAGGGTAGACTACCTGTTACTACATATTTGGCAAAAC
ATGGCAATGAGGAAAAGGTATATGAATTTTTAAAAAATGAACTTGGAAAAGGACATCAGGTTTATTTTGTATATCCTTTAATATCA
TCTTCAGAAAAGTTTAATCTAAAGGATGCTACTAGTATGTGTTTAAATCTTCAAAATATTTTTGTTGAATATTCTGTTGCTATGATT
CATTCTAAGCTTGAATCTTATGTGAAGGAAGAAATTATGCATAATTTTTATTTAAAGAAAATTGATATCCTGGTTGCAACGAGTGG
TAATGACCCTTTGAAGTCTTTAACTAAAGGGAAGGTATCTAGGGCTGGGAGAGATTCTTCTGGAAGAATTAGTGTTAGGAGAAGA

GGTGGAGGACATAAGAGACGTTATAGGGAAATTGATTTTGCCAGAAGGGATAAATTTGGTATACCTGCTAGAGTTGCAGCTATTG
AGTACGACCCAAATAGAAGCCCTAATATAGCTTTGCTTGTATATAGAGATGGAGATAAAAGGTATATTATTGCTCCAAAAGGAGT
TAAGGTTGGTGATATTTTGGAGAGTGGTCCAAATGCCCCAATAAAGATTGGTAATGCATTGCCGCTTGAAAATATTCCTGTTGGTA
AGATGGTTCACAACATTGAACTTAATATTGGAAAGGGTGGACAGCTTGTAAGAGGTGCTGGTGGATATGCTATGATTATTGCTTCT
GATGGAGATTATGTAACAGTTAAGCTTCCATCAGGTGAAATGAGAATGATTTTTAAGAAGTGCATAGCGACTCTTGGAGAAGTTG
GTAATGAAGATTATATGAATGTTTCTATTGGTAAAGCTGGTAAGAGTAGGTGGCTTGGTAGAAGACCTAAAGTTAGAGGTGTTGC
CATGAATCCTATAGATCATCCACATTCTGGAACTCTTTCAGGTGGTGAAGCTCAAAGAATTAGACTTGCTACACAAATAGGTTCAG
CTCTTGCTGGAGTTCTTTATGTACTTGATGAACCTAGCATTGGGTTGCATCAAAGAGATAATGAAAAATTAGTCAAGACCCTTGTT
AATCTTAAAGAACTTGGCAATACAGTAATTGTTGTAGAACATGATGAGCAAACTTTACGTACTGCTGATTATATTATTGATGTTGG
TCCTGGAGCTGGAATTTATGGTGGAGAGATCGTTGCTAAGGGAGTATTATCTGATATTTTAAATAATGAAAATAGTTTAACTGGTA
AATATTTAAGTGGACAACTTAAAATAGAAATTCCAAGAGTAAGACGTCAAATAGGAAAATCCAAAATTGTACTTTTGAAGGCAAA
TAAAAATAATTTAAAAAATATTAATGTTCACATTCCTTTGGGAGTTTTTACTGTAATAACTGGGGTTTCTGGTAGTGGAAAGAGTA
CTCTTCTTAATGAGGTATTATATCCTGCTTTGGATAGTAGGTTAAAAGCGAATGAAATCTATTCTGATGGTTTTGAGGAT

>Bhermsii
AGAGAAAGCATGATAATTGATGAGGAGATTTTATCTGATTTAATATTACACATTAAACTTTTAAGGATTAAATTCCTCTTTAAAAA
TAACACTCTTGGTATTTTTATTTTGATGGGTTCTTCTGATGTGGATAAGAATAAGCTTTCAGGTATTTTGTCAGAAGAACTTAAGAT
ACCCAAATTGACTTTGGGAGTGAGTGAGTATGGTGATTTTGATGGCATTAATCGATTAATAGGGCCTGTATATGGATCTGAATCTT
ATGATGAATTTACTACATTTTTTAAATTTTAAGTAAATCTTCAAGTTCAATTATTTTTCTATCCGATTTTGATAAATCTTCCAAGAG
AGTTATAGAATTTTTTTTTGAAGGATTTAATACGGGCAGACTTTTTGATAGTCTTGGTAGGAGTGTAAGCTTATCAGATAGTATAA
TACTAATAGATATTAATATAGAATATAGAGAACTTGCTAGTATTGGATTTAAAAATGAAACACTA---------
AATGGCAGGAGTTTGTTAGAAAAACGTTTTTCTAGTCAGTTTTTGGATCTTGTGGATCATATTTTTTTCTTTGCTGGATATGTTGGA
GAGGATGTTGAGAATATTTTACTTAAGTTGATTCATGCTGCTAATGGAGATGTGAGTTTTGCAGAAAGAGGTATTGTTTACATAGA
TGAAATAGATAAGATTGCAAAAAAAGGTGAAAATGTTTCAATTACAAGGGATGTATCTGGAGAAGGTGTGCAACAATCTTTATTG
AAAATAATTGAAGGCACTATTGCAAATATACCTCCAAAGGGTGGGAGAAAACATCCTTATGAGGAGACTATTGAAATCAATACGC
AAGACATACTATTCATATGTGGTGGGGCGTTTGTGGGACTTGAAAATATTATTAAAAGAAGAATTAATAAAAGTTCCATTGGCTTT
TCATCTGTTGGTAAACAAGATTCAAGGGAAAATAATTCTTTAAAATACTTAGAAATGGAAGATTTGGTTAAATTTGGACTCATTCC
AGAATTTGTCGGTAGGCTTCCTGTACATTCTTATCTTGAGAAGCTAGAAAAGGAAGATTTAATTAAAATATTAATTGAACCTGAGA
ATTCTATCGTGAAACAATATTATCATATGTTTAAAATGGATAATGTTGAACTGGTATTTGAAAGGGATGCGCTTAATGCAATTGCA
GAAGAAGCCATAAATGATAAAAAAATCATTTATTTTGATAATGCCGCAACTTCTCAAAAACCACAAAGCGTGATTTCATCTGTAG
TTGAATACTACGCAAACTATAATGCGAATGTTCACAGAAGCGGGCATGGACTTGCAATACAATCCAGCCTAAAAGTAGAAGAAA
CAAGAAAAATTATCAAAAAATTTATCAATGCAAAGTCTTTTAAAAACATAATATTTAATTCTGGAACTACAGATGGAACAAATAC
AATAGCAAATTCATTGCTCTTTTCAAAACTTTTAAAAGAAAATGATGAAATCATTACAACAACACTTGAACATAATAGTAATTTGC
TTCCTTGGATCAATATTGCTAAATATTTAAACTTAAAAATCAAACTTGCCAAATTTAATAAAATGGGAATCATTCAACCTGATGAA
ATAAAAAACTTAATTACAGATAAAACAAAAATTATTACTATATCTGGCATAAATAACATATTAGGAACTATACAGGATTTAGAAA
CGATTGGAAAGATTGCTAGAGATAACAAAATTATTCTGTTCATTGATGCAGCTCAAATGGCATATGATACTCATGAGAACATTAT
GATCATCACAAGTTTTAAAAAAAGCATTAAAGAAAAAATTATAGAAAAACTGCAAATATCTGAAAAAGATTTCTTGTCATGCGAT
TTAATATTCACAGCATCCGAACCTGCTAAAATTATAGGTAGCGAGGGCGAATTCTTAGCATCCAAAAATCTCGACAACAAATCAG
GATGTCATGCCATCATGAATGCATTTATTCATACAAATAACAATAAAAATAAAGCGGCTGTATTTTTTGATAATGAAGAGATCGG
ATCTGGAACTTCCAGAGGAGCTGATTCAAAACTATTAACAGAAATCTTAGAAAGAATTGATCATGTTCTAAATCTAGGGAAAGAA
GAACACTTAATTAAACTAAACAAATCATTCAATATCTCAATGGACGGTGCACATGGAACTCATCCAGGCTACATATGCAAACATG
ATCCAAATTATCAAATAAGTCTAGGAAAAGGCGTCACTATTAAAAGCAATGCTAATTTTAAATATGCAACAACAGCAAGTGGATG
TGCGAAACTTAAAGCTTTAGCTATGAAAAATAATATTGGGAGTGATAATATTGCTTTTGTGCACTTAACTTATATCCCAAATCCTG
CTGGGATAAATGAGCAAAAGTCTAAGCCTACTCAGCAGAGTGTCAAGACTTTAAATAAAGCTGGCATTTTTCCAGATTTAATTATT
GCAAGAAGTTCTCAACTTTTGACAAAGCAGATAAGACAAAAAATAGCAATGTTTTGTAATGTTGATGCTGAGTCCATTGTTGATA
ATACTGATGTTTCAACTATTTATGAAATACCCATATCTTTTTATAAACAGGGATTACACGAAATTTTGGGTTCAAAATTAAAAATA
AATGTT---
AAACCTAAGGTGGATGGGCTTGAGCGATTAGTAGGTATAATAAAGAGAAATTTTGCCTCTCCTCGTAAAGTTGTCAATATTGCTAT
TTGTGGTAAATATACTGAGCTTGGTGATTCTTATGCATCAATATTTGAGTCTTTAACCCATGTATCTGCTAATTTAGATCTTTTAGT
TAAGACAATTGTAATTGATAGCACTAATTTTGATGAAGGGATGTTAAAGGGTATAGATGGTATAGTAGTGCCTGGTGGATTTGGG
GGTAGGGGATATGAAGGTAAAATGCGTGCAATTGCGCGGCAACATTATAATAATTTAGCAAATATATTAAAAGATTTTAATGTTT
CTATAGCTTTTTTAACTGGTAGTTTAAAAAAGAGAGATAGAGATGAGGTTTTAGAGAAAATTCAAAGTGGAACTTATAGTTTAGT
AATTGGAACTCATGCTATTTTTTCTCAGGGAACAAAATTTAAAAGATTAGCTTATATCATTATTGATGAACAGCATAAATTTGGTG
TTGAGCAGAGAGAGGAGCTTAAAAATAAGGGAGAGGAGGTTGATGTGCTTTTAATCTCAGCAACTCCTATTCCTAGAAGCTTAGC
TTTGACCCTTTTTGGCGATCTTGAGATATCTTTAATTAGAAGAGGTCCTGCAGGTCGAATACCTGTTACTACGTATTTAGCAAAAC
ATGGCAATGAAGAGAAGGTATATGAATTTTTAAAAAATGAACTTGCAAAAGGACATCAGGTTTATTTTGTTTATCCCCTAATATCA
TCTTCACAGAGGTTTAATTTAAAAGATGCTACTAGTATGTGTTTAAATCTTAAGAATATTTTTGTTGAGTATTCTGTTGAAATGATT
CATTCTAAGCTTGAATCTCATGTTAAGGAAGAGATTATGCATGATTTTTACTTAAAGAGAATAGATATTTTAGTTGCAACGAGTGG
TAATAACCCTTTGAAGTCTTTAACTAGAGGGAAGGTATCTAGGGCTGGAAGAGATTCTTCTGGAAGAATTAGTGTTAGAAGAAGA
GGTGGAGGACATAAGAGACGGTATAGAGAGATTGACTTTGGCAGAAGGGATAAGTTCGGTATACCTGCTCGAGTTGTATCTATTG
AGTATGATCCAAATAGAAGTTCTAATATAGCTTTGCTTGTTTACAAAGATGGGGATAAGAGATATATTATTGCTCCTAAAGGAATT
AAAGTCGGCGATATTTTAGAGAGCGGTCCAAATTCCCCAATAAGGATTGGTAATGCTTTGCCACTTGAGAATATTCCTGTTGGTAA
GACAGTGCATAACATTGAGCTTAATATTGGAAAAGGTGGACAGCTTGTAAGAGGTGCTGGGGGATATGCGATGGTGCTTGCTTCT
GAGGGAGATTATGTAACAGTTAAGCTTCCATCAGGCGAGATGCGCATGATTTTTAAGAAATGCATGGCAACTCTTGGAGAAGTTG
GAAATGAAGATTACGTAAATGTTTCTATTGGTAAGGCTGGTAAGAGTAGATGGCTTGGCAGAAGGCCTAAAGTAAGAGGTGTTGC
GATGAATCCTGTAGACCATCCGCATTCTGGAACCCTTTCAGGAGGAGAGGCTCAACGAATTAGACTTGCTACTCAAATAGGTTCA
GCTCTTGCTGGGGTTCTTTATGTACTTGATGAGCCTAGTATTGGATTGCACCAAAGAGATAATGAGAAATTAATAAGTACACTTGT
TAATTTGAAAGAACTTGGCAATACGGTAATTGTTGTGGAGCATGATGAGCAAACTTTACGCACCGCTGACTATATTATCGATGTTG
GGCCTGGAGCTGGGATTTATGGTGGAGAGATAGTTGCTAAGGGAACCTTATCTGATATTTTAAGTAATGAGAATAGTTTGACTGG
TAAGTATTTAAGCGGTCAACTTAAAATAGAAGTTCCAAAAACAAGACGTAAGGCAGGAAAGCTGAAATTGTACTCTTGAATGCT
AATAAAAATAATTTAAAGAATATTAATGTGCGCATTCCTTTGGGAGTTTTTACTGTAATAACGGGAGTTTCTGGTAGTGGAAAGAG
TACCCTTCTTAATGAGGTATTATATCCTGCTCTTGATAGTAGGTTAAAATCCAATACAAGTTATTTTGATGGCTTTGAGGAT

>Bturicatae
AGAGAGAGTATGATAATTGATGAGAATATTTTATCTGATTTGATATTAAACATCAAACTTTTAAGGATTAAATTTCTCTTAAAAAA
TAATACTCTTGGTATCTTCATTTTGATGGGTTCTTCTGATGTAGATAAGAATAAGCTTTCATGTATTTTATCAGAAGAACTTAAGAT
ACCTAAATTGACTTTGGGAATGAGTGAATATGGTGATTTTGATGGCATTAATAGATTAATAGGACCTGTATATGGATCTGAATCTT
ATGATGAACCTACTAAGTTTTTTAAATTTTTAAGTGAATCTTCAAGTTCAATTATTGTCCTGTCTGATTTTGATAAATCTTCCAAGA
GGGTTATAGATTTTTTTTTTGAAGGATTTAATACGGGTAGGCTTTTTGATAGTCTTGGTAGAAGTGTAAGCTTATCAGATAGTATA
ATAATACTAGATATTGATATAAAACATAGGGAATTTGATGGTATTGGATTTAAAACCGAAACAATC---------
GATGGTAGAAGTTTGCTAGAGAAACGTTTTTCTACTCAGCTTTTGGATCTTGTGGATCATATTTTTTTCTTTGCTGGATATGTGGGA
GAAGATGTTGAGAATATTTTACTTAAGCTAATTCATGCTGCTGATGGAGATGTAAGTTTTGCAGAAAGGGGTATCATTTACATAGA
TGAAATAGATAAGATTGCAAAAAAAGGTGAAAATGTTTCAATTACAAGGGATGTATCTGGAGAAGGTGTTCAACAATCTTTATTA
AAAATAATTGAAGGCACTATTGCAAATGTACCGCCAAAAGGTGGGAGAAAGCATCCTTATGAGTGTACTATTGAAATCAATACGC
AGAACATACTATTTATATGTGGTGGGGCTTTTGTAGGACTTGAAAATATTATTAAAAAAAGAATTAATAAAAGTTCTATTGGATTT
TCATCTGCTGGTAGAAAAGATTCTGGAGAGATTAATTCTTTAAAATACTTGGAAATGGAAGATTTGGTTAAATTTGGACTCATACC

GGAGTTTGTAGGTAGGCTTCCTGTACATTCTTATCTTGAAAAGTTAGAAAAGGAAGACCTGATTAAAATATTAATTGAACCTGAA
AATTCTATCGTGAAACAATATTGTCATATGTTTAAGATGGATAATGTTGATCTATTATTTGAAAGAGATGCGCTTGATGCAATTGC
AGAAGAGGCCATTAATGATAAAAAAATCATTTATTTTGACAATGCCGCAACTTCTCAAAAACCACAAAGTGTGATTTCATCTATA
GTTGAATACTACACAAATTATAATGCAAATGTGTACAGAAGCGGACATGAACTTGCAATACAATCCAGCTTAAAAGTAGCAAAAG
TAAGGAAAATTGTCAAAAATTTTATCAATGCAGAATCTTGTAAAAACATAATATTTAATTCTGGTACTACAGATGGAATAAACAC
AGTAGTAAACTCATTACTTTTTTCAAAACTTTTAAAAAAAAATGACGAAATCATTACAACAACGCTCGAACACAATAGTAATTTGC
TTCCTTGGATAAATATTGCCAAATTTTTAAAATTAAAAATTAAACTTGCAAAATTTAATGAAATGGGAATCATTCAACCAGAACAA
ATAAAAAACTTAATTACAAATAAAACAAGAATTATTGCTATATCTAGCATAAACAATATACTAGGAACTATACAAGATTTAGAAA
CAATTGGAAAGCTTGCTAGAGATAATAAAATTATTTTATTCATAGATGCAGCTCAAATGGCATATGATACTCATGAGAACATAGT
CATCATCACAAGTTTTAAAAAAAGCATTAAAGAAAAAATTCTAGAAAAACTGCAAATATCTGAAAAAGATTTTCTATCATGCGAT
TTAATATTTACAGCATCAGAACCGGCTAAAATTATAGGTAGCGAAGGTGAATTTTTAGCATCCAAAAACCTAGATAACAAATCGG
GATGCCATGCCATCATGAATGCATTTGTTCATACAAATAACAATAAAAATAAAGTAGCTGTATTTTTTGACAATGAAGAGATTGG
ATCTTTAACTTCCAGAGGAGCTGACTCAACACTATTAACAGAAGTCTTAGAAAGAATTGATCATGTTTTAAATCTAGGAAAAGAA
GAACATCTAATTAAACTAAACAAATCATTCAATATTTCAATGGACGGCGCACATGGGGTTCATCCAGGCTATATATGTAAACATG
ACCCAAACTATCAAATAAGTCTAGGGAAAGGAGTAACCATCAAAAACAATGCCAATTTTAAATATTCAACAACAGCACATGGAT
GTGCAAAACTTAAAGCCTTAGCTATGAAAAATAATATTGGGAGTGATAATATTGCTTTTGTGCACTTAACTTATATTCCAAGTCCT
GTTGGAATAAATGAGCAAAAGTCTAAGCCTACTCAGCAGAGTGTAAAGACTTTAAATAAAGCTGGAATTTTTCCAGATTTAATTA
TCGCAAGAAGTTCTCAACTCTTAACAAAACAGATGAGACAAAAAATAGCAATGTTTTGTAATGTTGATACTTTATCCATTATTGAT
AATACTGATGTTTCAACTATTTATGAAATACCCATATCTTTTTATAAACAGGGATTACACGAAATTTTGGGTTCAAAGTTAAAAAT
AAATGTT---
AAGCCTAAGGTGGATGGGCTTGAGCGATTAGTAGGTATAATAAAGAGGAATCTTGTTTCTCCTCAAAGGATTGTAAATATTGCTA
TTTGTGGTAAATATACTGAGCTTGGTGATTCTTATGCATCAATATTTGAGTCTTTAACCCATGTGTCTGCTAATTTAGATATTTTAG
TTAAAACAACTGTAATTGATAGTACTAATTTTGATGAAAGGATGTTAAAGGGTATAGATGGTATAGTAGTGCCTGGTGGATTTGG
GGGGAGAGGATATGCAGGTAAAATTCGTGCAATTGCACGGCAACATTATAATAATCTATCAAATATGTTAAAAGATTTTAATATT
TCAATAGCTCTTTTGACTGGTAGTTTGAAAACGAAAAATAGAAATGATGTTTTAGAAAAAATTCAAAGTGGCACTTATAGTTTAGT
AATTGGGACTCATGCTATCTTTTCTCAAAGAACAAAATTTAAAAAATTAGCTTATGTTATTATTGATGAACAGCATAAATTTGGTG
TTGAGCAGAGAGAGAAGCTTAAAAATAAGGGAGAAGAAGTTGATATGCTTTTAATGTCAGCAACCCCTATTCCTAGAAGCTTAGC
TTTGACTCTCTTTGGTGATCTTGAGGTATCTTTAATTAAGAGAGGTCCTGCAGGTCGAAGACCTGTTACTACTTATTTAGCGAAGC
ATGGAAATGAGGATAAGGTATATGAATTTTTAAACAATGAGCTTGGAAAAGGACATCAGGTTTATTTTGTTTATCCTTTAATATCA
TCTTCAGAGAAGTTTAATTTAAAAGATGTTACTAGTATGTGTTTAAATCTTAAGAATATTTTTGTTGAATATTCTGTTGCAATGATT
CATTCTAAGCTTGAATCTCATGTTAAAGAAGAAATTATGCGTGATTTTTATTTGAAAAAAATAGATATTTTAGTTTCAACAAGTGG
TAATAGTCCTTTGAAGTCTTTAACTAGAGGGAAGGTATCTAGGGCTGGAAGAGATTCTTCTGGAAGAATTAGTGTTAGAAGAAGA
GGTGGAGGACATAAAAGACGGTATAGAGAGATTGACTTTGGCAGAAGGGATAAGTTTGGTATACCTGCTCGAGTTGTGTCTATTG
AGTATGATCCAAATAGAAGTTCTAATATAGCTTTGCTTGTTTACAGAGATGGGGATAAAAGATATATTATTGCTCCTAAAGGAATT
AAAGTCGGTGATATTTTGGAGAGTGGTCCAAATGCTCCAATAAGGATTGGCAATGCTTTGCCACTTGAGGACATTCCTGTTGGTAA
AGCAGTGCATAACATTGAGCTTAATGTTGGAAAAGGTGGACAGCTTGTAAGAGGTGCTGGAGGCTATGCGATGGTGCTTGCTTCT
AAGGGAGATTATGTAACAGTTAAGCTTCCATCAGGTGAGATGCGAATGATCTTTAAAAAATGCATAGCGACTCTTGGAGAAGTTG
GCAATGAAGATTATATAAATGTTTCTATTGGTAAAGCTGGTAAGAGTAGATGGCTTGGTAAAAGGCCTAAGGTAAGAGGTGTTGC
GATGAATCCTATCGACCATCCACATTCTGGAACTCTTTCAGGGGGTGAGGCTCAACGAATTAGACTTGCTACTCAAATAGGCTCGG
CTCTTGCTGGTGTTCTTTATGTACTTGATGAGCCTAGTATCGGATTACATCAAAGGGATAATGAAAAATTAATAAGTACTCTTGTT
AATTTGAAAGAGCTTGGTAATACGGTAATTGTTGTAGAGCATGATGAGCAGACTTTGCGTACTGCTGATTATATTATTGATGTTGG
ACCTGGAGCTGGGATTTATGGTGGAGAAATAGTTGCTAAGGGATCTTTATCTGATATTTTAAATAATGATTATAGTGTGACTGGAA
AGTATTTAAGTGGTCAACTTAAAATAGAAGTTCCAAAAATAAGACGTAAAGTAGGAAAAGCTGAAATTGTACTCTTAAATGCTAG
TAAAAATAATTTAAAAAATATTAATGTGCGTATTCCTTTGGGAGTTTTTACTGTAATAACAGGAGTTTCTGGTAGTGGAAAGAGTA
CCCTTCTTAATGAGATATTATATCCTGCTCTTGATAGTAGGTTAAAATCAAATACAAGTTATTTTGATGGTTTTGGGGAT